# UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FILOSOFÍA

Departamento de Teoría del Conocimiento, Estética e Historia del Pensamiento



## TESIS DOCTORAL

# El problema de la conciencia en la filosofía de la mente y las ciencias cognitivas: replanteamiento del núcleo del debate y valoración crítica de los principales marcos teóricos

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR


**Asier Arias Domínguez**

Directores

**Pedro Chacón Fuertes**
**José Antonio Guerrero del Amo**


**Madrid, 2016**

# UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FILOSOFÍA

Departamento de Teoría del Conocimiento, Estética e Historia del Pensamiento


## TESIS DOCTORAL

**El problema de la conciencia en la filosofía de la mente y las ciencias cognitivas. Replanteamiento del núcleo del debate y valoración crítica de los principales marcos teóricos**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Asier Arias Domínguez**

Director

Pedro Chacón Fuertes

Codirector

José Antonio Guerrero del Amo

**Madrid, 2015**

Universidad Complutense de Madrid
Facultad de Filosofía
Departamento de Teoría del Conocimiento, Estética e Historia del Pensamiento.

# EL PROBLEMA DE LA CONCIENCIA EN LA FILOSOFÍA DE LA MENTE Y LAS CIENCIAS COGNITIVAS.

## Replanteamiento del núcleo del debate y valoración crítica de los principales marcos teóricos

Memoria presentada para optar al Grado de Doctor por Asier Arias Domínguez bajo la dirección de los Doctores Pedro Chacón Fuertes y José Antonio Guerrero del Amo.
Madrid, 2015



Tesis defendida en noviembre de 2015. Revisada en septiembre de 2016

# ÍNDICE

## PARTE TERCERA

Crítica de los principales marcos teóricos

*A la memoria de Avelino Arias Pérez y Bienvenida Hernández Lázaro, mis abuelos paternos.*

«La conciencia es la pesadilla de la naturaleza.»

EMIL M. CIORAN

*De lágrimas y de santos*

«Para que vuelva a revelársete la vida con toda su magnificencia torna la vista hacia el sol, y despréndete al despertar de entre las alas de tu débil sueño.»

JOHANN W. GOETHE

*Fausto*

**Abstract**

The problem of consciousness is at the very core of contemporary philosophy of mind and cognitive sciences. It is quite complicated to take three consecutive steps in any given subarea of cognitive sciences without coming across this problem in one way or another. Nevertheless, what is the problem of consciousness? We could say it is the problem of explaining how the overall organized activity of something like a nervous system can give rise to something like a pain or joy experience. Of course, this is just a caricature: there are countless open debates in many specialized topics that fall under the so-called "problem of consciousness", so much so that some pundits have been speaking for two decades of an interdisciplinary project called *Consciousness Studies* that would bring together specialists in philosophy, psychology, neurosciences, artificial intelligence and related disciplines. The following pages are aimed, on the one hand, at providing an overall perspective of this interdisciplinary field of research and, on the other, at refuting those who have been proclaiming the impossibility of solving the problem of consciousness –suggesting, furthermore, feasible pathways towards a solution.

The thesis is divided into three main parts: the first one is eminently expositive, the second argumentative and, the third, critical. We have entitled the first part "Historical and conceptual approach to the problem of consciousness in contemporary philosophy and science". Arguably, the heading does not seem to require clarification or justification: we will start at the beginning, namely identifying the problem and briefly presenting the state of the art in the relevant disciplines. In order to achieve this goal, Chapter 1 introduces the problem from a historical point of view to provide the pertinent context. In Chapter 2 we shed some light on the plural nature of the problem: as we shall see, there is not one single problem of consciousness, nor a single form of consciousness. Chapter 3 addresses one of the most controversial issues on *Consciousness Studies*: the phenomenal/intentional duality of the mental. In the first part of the thesis, this topic will be raised on a purely expository basis: we will define the intentional and the phenomenal, thus laying the foundations for the argument to be developed in the second part about the relations between these supposedly exhaustive, specific, homogeneous and precisely delimitated aspects of the mental. Meanwhile, Chapter 4 outlines a schema of the ontological conceptions of consciousness, whereas Chapter 5 does the

same for the different explanatory proposals. Finally, in Chapter 6 we will deal with the arguments intended to demonstrate that consciousness is an unapproachable object of study for the natural sciences. Such arguments seek to bolster the inexplicability of consciousness intuition, a customary hunch according to which any given theory of consciousness will collide with an insurmountable impasse. Once we reveal the low strength of these arguments, we will reformulate, in the second part, the terms the problem of consciousness is currently examined under.

After defending in the first part that there is no motive for discouragement and that there are no reasons of principle for supposing the inexplicability of consciousness, we turn to a "rethinking of the contemporary debate on the problem of consciousness" –the heading of the second part. To this end, in Chapter 7 we put forward a succinct defence of the naturalistic base from which we delve into our rethinking. Then, in Chapter 8, we argue that the problem of consciousness is the one the explanatory theories face and, hence, the ontological debate can not contribute to its solution. Having explicitly defined the problem and the different means to approach it, we take into consideration how it is usually treated, arguing that the endemic philosophical speculation about the relationship between the abstractions "intentional mind" and "phenomenal mind" can hardly help explain consciousness, and rejecting the oversimplification of the interparadigmatic struggle between advocates of representationalism, inseparatism, enactivism or any other attempt to convince the scientific community that consciousness is a singular phenomena to be explained from a single theoretical framework or by a single discipline. In Chapter 9 we outline a rough sketch of a theoretical groundwork for the biology of consciousness, stressing the need to transcend the intellectualist tradition –which considers the mental as cognitive, the cognitive as calculation, and consciousness as the pinnacle of the cognitive pyramid– and taking seriously into account theoretical and experimental tools from affective neurosciences and comparative psychobiology. Finally, in the third part we comment the theoretical frameworks of Searle, Dennett and the neurophenomenologists.

**Materiales**

Aunque una notable cantidad de aficionados a la cita de lance ha venido atribuyéndosela a Oscar Wilde o a George Bernard Shaw, fue en verdad William Somerset Maugham quien, en la segunda edición de *The Creative Impulse*, de 1931, describiera a uno de sus personajes haciendo uso de esa conocida frase según la cual la pericia en el arte de la cita es un solvente sustituto del ingenio. Como no podía ser de otro modo, esta tesis está plagada de citas. Citamos, en cualquier caso, siguiendo las versiones originales de los trabajos utilizados, esto es, citamos en la lengua en la que fueron escritos, aunque en ocasiones traducimos determinados pasajes y en ocasiones aludimos a determinadas traducciones, haciendo lo primero cuando entendemos que el texto y el contexto así lo piden,[1] y lo segundo, bien con la intención de restituir al original lo que una traducción inatenta hubiera podido detraerle, bien con la de, sencillamente, facilitar al lector el acceso al pasaje de que se trate en la traducción que sea el caso. En la bibliografía incluimos siempre, entre corchetes y a continuación de la original, la referencia de la traducción castellana disponible, hayamos o no aludido a ella en el texto.

Por otra parte, algunas de las ideas expuestas en determinados lugares de esta tesis fueron previamente presentadas en congresos, seminarios y talleres dentro y fuera de nuestras fronteras. Son las mismas, en este sentido, deudoras del intercambio con un número de especialistas profuso al punto que corresponderles listando sus nombres forzaría al lector a una escasamente feraz lectura diagonal y a nosotros a un ejercicio mnemónico de resultas inexcusablemente parciales. Listamos a continuación, eso sí, los señalados eventos en orden cronológico e incluyendo sucintos comentarios acerca de aquello que de cada uno de los mismos afluyera a esta tesis.

En el IX Congreso Internacional de Antropología Filosófica presentamos la comunicación "El problema mente/cuerpo a partir de la neurociencia cognitiva", cuyo texto apareciera posteriormente publicado en *Thémata. Revista de Filosofía* bajo el título "Del problema mente/cuerpo al estudio de la mente-cerebro. A partir de la neurociencia cognitiva". De la reflexión que allí emprendiéramos acerca de las convergencias y divergencias de las concepciones de lo mental defendidas por Michael Gazzaniga, Ge-

[1] Indicamos en este caso, incluyendo junto a la entrada a la referencia la letra griega tau (τ), que la traducción es nuestra.

rald Edelman y el neuroconexionismo de La Jolla proviene la orientación de nuestra exposición de las mismas en el capítulo quinto.

En el V Congreso Internacional de la Sociedad Académica de Filosofía presentamos la comunicación, publicada en actas, "La naturalización de la razón. ¿El último bastión de la resistencia al programa naturalista?", algunos de cuyos planteamientos recoge el primer capítulo de la segunda parte.

En las XLVIII Reuniones Filosóficas de la Universidad de Navarra presentamos la comunicación "El marco filosófico de la neurobiología de la conciencia de Gerald Edelman". Esta ponencia y, asimismo, el esquemático artículo de divulgación publicado en la revista *Ciencia Cognitiva* bajo el título "Neurociencia de la conciencia. Introducción al marco teórico de un clásico contemporáneo" constituirían el substrato de nuestra aproximación a la teoría de la conciencia del biólogo neoyorquino.

En la Retecog Summer School 2012, *Cognition & Consciousness*, en la que participaran Anil Seth y Ap Dijksterhuis, presentamos la comunicación "Mysterianisms and naturalisms in contemporary Consciousness Studies: The explanatory gap", de la que partiría la crítica del argumento de la brecha explicativa que desarrollamos en el capítulo sexto.

En la XVI Consciousness and Experiential Psychology Section of the British Psychological Society Annual Conference, en la que participara Max Velmans, presentamos la comunicación "Neurophenomenology, first/third-person methodologies, and deflationism", de la que provienen algunas de las acotaciones críticas propuestas en el tercer capítulo de la tercera parte.

En la I International Krakow Conference in Cognitive Science, en la que participaran Thomas Metzinger y David Papineau –que realizara un enriquecedor comentario a nuestra ponencia–, presentamos la comunicación "Consciousness and intentionality: Troubles with inseparatism", a partir de la cual comenzáramos a elaborar la crítica del inseparatismo searleano que puede leerse en el segundo capítulo de la segunda parte.

Fuimos amablemente invitados a impartir la sesión XIV del Seminario Interuniversitario de Filosofía de la Ciencia de Madrid, que tituláramos "La neurofenomenología como metodología para una ciencia de la conciencia". De las ideas que en dicha sesión expusiéramos derivan algunas de las perspectivas críticas con que cerramos el tercer capítulo de la tercera parte.

En el V Congreso Internacional de Jóvenes Investigadores en Filosofía presentamos la comunicación "Falencias ontológicas del naturalismo biológico searleano.

Emergentismo causal antirreduccionista: oyendo misa y repicando", de la que proceden algunas de las ideas que desarrollamos en el primer capítulo de la tercera parte.

En el IV Encuentro de Estudiantes del Doctorado Interuniversitario en Lógica y Filosofía de la Ciencia presentamos, en calidad de ponente invitado externo al Programa Interuniversitario de Doctorado en Lógica y Filosofía de la Ciencia, la comunicación "El problema de la conciencia en John R. Searle. ¿Naturalismo o dualismo biológico?", en la que discutimos asimismo algunos de los problemas de la ontología searlena de los que tratamos en el referido capítulo primero de la tercera parte.

Apuntemos para terminar que nuestra exposición de la revolución cognitiva en el primer capítulo saca partido, por más que desde cierta distancia, de los materiales utilizados en la preparación de los artículos: "La neurociencia computacional en Churchland & Sejnowski", publicado en la revista *A Parte Rei*; "Avatares del paradigma conexionista", publicado en la revista *Ciencia Cognitiva*; y "Fundamentos de inteligencia artificial. Entrevista con Antonio Benítez", publicado en la misma revista. Por su parte, nuestra conceptualización de la noción de *qualia* lo hace de los utilizados en la preparación de los artículos: "¿Ha logrado Dennett quinear los qualia? Una revisión naturalista", publicado en la revista *A Parte Rei*; "A qué se refieren los filósofos de la mente contemporáneos cuando dicen *'qualia'*", publicado en la revista *Apeiron*; y "Los qualia: Intuiciones y argumentos. Apuntes para una nueva aproximación", publicado en la revista *Cuaderno de Materiales*. Por último, el exordio que subsigue a esta sección lo hace de los utilizados en la preparación del texto divulgativo "La conciencia: el truco más sorprendente del ilusionista más inescrutable", publicado en *Entretanto Magazine*.

## Exordio

Despierto de un sueño sin ensoñaciones. La actividad alfa y beta ampliamente distribuida que un equipo electroencefalográfico podría registrar a través de mi cráneo vuelve a sustituir a la regular y sincrónica actividad delta y los husos del sueño, y la ausencia, el silencio y la oscuridad de un no-mundo en el que no soy quedan nuevamente atrás, como un recuerdo vago, evanescente, un recuerdo que, de hecho, no puedo tener porque es acerca de nada y se extiende a lo largo de un tiempo que, por lo que a mí respecta, no transcurrió, y vuelvo así a la multimodal pirotecnia technicolor de la experiencia consciente, a sentir y sentirme. El problema de la conciencia es precisamente éste: un interruptor hipotalámico induce cambios en el equilibrio neuroquímico de las áreas encefálicas relacionadas con la modulación del arousal –como los núcleos del rafe o el locus coeruleus– y un mundo se abre o se cierra para el organismo; un a día de hoy indeterminado conglomerado de procesos biológicos acaece y, como por arte de magia, ese conglomerado, observable desde fuera, puede también observar desde dentro; esos procesos, pasibles de una hipotéticamente completa descripción en tercera persona que no hiciera mención de ello, pueden ahora, sorprendentemente, dar una descripción de sí en primera persona: aparece un sujeto, un punto de vista.

¿Y cómo encajan un "sujeto" o un "punto de vista" en un universo como el descrito por la física, un universo compuesto, exclusivamente, por puntos de masa-energía inmersos en campos de fuerzas? Una pelota encaja fácilmente en un universo como ése. La dolorosa sensación subjetiva que causa un pelotazo en la zona cuadril, en cambio, es un poco más difícil imaginar cómo podría encajar en semejante universo porque, para empezar, ¿puede una sensación expresarse en términos de magnitudes físicas? Parece que una experiencia consciente ni puede pesarse, ni determinarse a qué presión o temperatura se encuentra, ni cuál es su volumen, ni hasta qué punto es dura, plástica o resiliente, ni tampoco cuáles son las propiedades de los campos físicos asociados a ella. Así, intuitivamente, da la impresión de que los estados mentales conscientes no comparten ninguna propiedad con el mundo descrito por la física. La reacción tradicional a esta intuición de incompatibilidad consistió en postular ontologías dualistas en las que la naturaleza de los fenómenos conscientes era presentada como irreconciliable con la del mundo descrito por la física. Según esta reacción dualista, pues, los fenómenos conscientes son cosa del alma, de la que nada puede decirnos la física moderna.

La reacción dualista puede resultar comprensible en vista del peculiar carácter de la conciencia como objeto de estudio científico, y no sólo desde el punto de vista de la física, sino desde el de las ciencias naturales en general. Así, por ejemplo, según la teoría de la evolución la existencia de la inmensa mayoría de los rasgos biológicos responde –o respondió– a un incremento de la aptitud biológica de sus portadores, esto es, a una aumentada capacidad para engendrar dado el mejorado desempeño de la función biológica que el rasgo en cuestión contribuya a realizar. Pero, ¿qué función biológica cumple la conciencia? ¿Cómo podría ella aumentar la aptitud biológica de un organismo? Si la conciencia no cumpliera ninguna función biológica figuraría como *rara avis* en el catálogo de los rasgos biológicos, contaría como injustificada excepción: habría evolucionado *porque sí*. Y lo curioso es que, al detenernos a considerar qué función biológica podría desempeñar la conciencia en la economía de un organismo surgen razones que sugieren que ninguna. Veámoslo haciendo uso de algunos ejemplos. Sabemos que determinadas conductas pueden realizarse en ausencia de experiencia consciente. La vía visual dorsal puede hacer que nuestro párpado proteja nuestro ojo ante la acometida de una astilla aunque no la percibamos conscientemente; se ha informado de sujetos que, en medio de crisis epilépticas tipo *petit mal*, caminan hasta llegar a sus casas o siguen interpretando al piano la pieza que estaban tocando mientras eran tan conscientes como la silla en la que estaban sentados; pacientes con ceguera histérica pueden esquivar objetos que dicen no ver. Así las cosas, ¿cómo vendría la conciencia a incrementar la aptitud biológica de un organismo cuando parece que éste, como sugieren los ejemplos aducidos, podría conducirse y habérselas sin ella? De nuevo nos vemos intuitivamente impelidos a la reacción dualista: la conciencia aparece como algo excepcional en el mundo natural y no sabemos muy bien cómo podrían las ciencias naturales incluirla entre sus objetos de estudio.

Es bien sabido, sin embargo, que la reacción dualista no es una opción exenta de problemas teóricos fundamentales. Según esta reacción tradicional existirían, por una parte, entidades y fenómenos físicos y, por otra, entidades y fenómenos mentales, que residirían en el alma y serían, como ella, de naturaleza inmaterial. Pero, ¿cómo podría mi alma inmaterial mover a voluntad mi brazo material? ¿Cómo interactuarían estas dos esferas ontológicas irreconciliables? Los esfuerzos de los dualistas por responder a esta pregunta sin producir mayor desconcierto del contenido en el interrogante han sido vanos y han hecho, además, evidente el modo en que el dualismo alma/cuerpo viola ineluctablemente las leyes fundamentales que, según parece, rigen el universo: cada vez

que un alma inmaterial mueve un brazo material, la ley de la conservación de energía, una ley para la cual, como dijera Richard Feynman, "no hay excepción conocida" (Feynman, 1963: 67 de la traducción), es violada por la introducción de energía cinética en un universo desconcertado ante los constantes insumos de terajoules procedentes de ninguna parte y canalizados por trillones de almas capaces de mover, a voluntad, los cuerpos que habitan aquí y allá, en la tercera roca desde el Sol o en algún lugar de la Galaxia del Cigarro (a.k.a. Meiser 82). Pero la reacción dualista no resulta problemática por sus consecuencias ontológicas, esto es, por la imagen que ofrece de la estructura del universo. La más fidedigna entre las versiones disponibles de esa imagen procede de los medios epistémicos acreditados (los métodos y teorías científicas), y el verdadero problema de la reacción dualista es su inoperancia y munificencia respecto de esos medios, ambas perfectas, bien que la primera en su exceso y la segunda en su defecto.

El principal propósito tras las páginas que subsiguen es el de poner de relieve que cabe esperar que continuar planteando la cuestión en términos ontológicos sirva a la solución del problema de la conciencia en la misma medida en que ha venido haciéndolo. Mientras tanto, sigue abierta la posibilidad de estimular el progresivo refinamiento de nuestras prácticas epistémicas, dejando que sean las resultas de éstas las que, como hasta ahora, paulatina y eventualmente iluminen nuestra ontología.

## Introducción

Contra la norma, vamos a ocuparnos del problema de la conciencia sin la intención la de elaborar una "nueva teoría de la conciencia" –quizá la más manida entre las locuciones hiperbólicas de uso frecuente en los *Consciousness Studies*–. Nuestro primer objetivo será el de definir el problema. A ello dedicaremos la práctica totalidad de la primera parte, al cabo de la cual debieran hacerse ya evidentes los motivos por los cuales consideramos oportuno ahorrarnos una "nueva teoría de la conciencia", unos motivos en los que profundizaremos en la segunda parte y que cabe extractar apuntando que la conciencia es un fenómeno enormemente vasto cuyo estudio requiere –y, con toda seguridad, seguirá requiriendo, bien cabe que indefinidamente– la integración de una considerable cantidad de disciplinas y herramientas teóricas antes que un nuevo ejercicio especulativo unilateral y pretendidamente fundamental, unificador y omniexplicativo.

No propondremos una "nueva teoría de la conciencia", pero patentizaremos que el de cantar la palinodia es el destino que se ajusta a los confalonieros de la doctrina misteriana. Esta doctrina se alza sobre una tesis según la cual el proyecto de explicar científicamente la conciencia está condenado al fracaso, una tesis con un sustento equiparable al de ésa otra según la cual si acaso nos atreviéramos a lanzar un dado no saldría un cuatro. A desmontar esta tesis dedicaremos un capítulo completo, el que cierra la primera parte. El objeto de la segunda, por su parte, será el de contribuir a allanar el camino hacia esa explicación imposible según el misteriano. Hasta el momento, ese camino ha tratado de abrirse en dos direcciones: la ontológica y la explicativa. La distinción, aunque en cierta medida artificial, sirve para diferenciar entre los intentos de alcanzar una adecuada comprensión de qué sea la conciencia y cuál haya por tanto de ser el lugar que ocupe en nuestra concepción de la realidad, por una parte, y los destinados a esclarecer las causas y mecanismos de su existencia, por otra. Argumentaremos que seguir deambulando por la vía ontológica difícilmente podrá favorecer nuestra comprensión de la naturaleza de la conciencia. No es preciso, pues, continuar echando leña al fuego de las etiquetas barrocas, los conceptos inoperacionalizables y las hipótesis inconfutables, sino cooperar en la investigación del modo en que esos fenómenos a los que denominamos conscientes surgen en la historia evolutiva y acaecen actualmente. Esta última frase contiene de forma implícita el fondo y la médula de cuanto esta tesis aspira a ofrecer por

cuanto sugiere que el problema de la conciencia a) puede resolverse y b) que ello no depende de la especulación de carácter metafísico, sino de la investigación de su fisiología, su etología y su filogénesis. Así pues, dividiremos el problema de la conciencia mediante una línea imaginaria que dejaría a un lado su aspecto funcional (¿de qué modo la actividad de un organismo en interacción con su medio origina experiencia consciente?) y al otro su aspecto histórico (¿cómo surgieron en la filogénesis los organismos conscientes?). En cuanto al primero de estos aspectos, no nos detendremos a estimar el mérito relativo de los planteamientos teóricos que del mismo han tratado de dar cuenta, sino sólo su plausibilidad filogenética y su utilidad de cara a esclarecer la historia evolutiva de la conciencia. En lo que a ésta se refiere, destacaremos la a menudo desatendida necesidad de integrar la explicación de la conciencia en un contexto evolutivo y apuntaremos a una serie de restricciones ineludibles de cara a cimentar un marco teórico capaz de encarnar la amplitud, complejidad y flexibilidad impuesta por la interdisciplinariedad que a su vez nuestro objeto sanciona.

En vista de la comentada amplitud del campo de estudio que los *Consciousness Studies* se proponen cubrir en su intento de dar cuenta de, quizá, el más extenso, heterogéneo, plural y complejo entre todos los conjuntos de fenómenos biológicos, la alternativa razonable al acostumbrado engendramiento de teorías especulativas, unitarias, unilaterales, fundamentales y reduccionistas es la de abrir vías hacia la integración y la convergencia de una gran cantidad y variedad de programas de investigación circunscritos a diferentes niveles de análisis. A nadie debiera sorprender, en otras palabras, que no fuera una teoría fundamental, una *teoría cognitiva del todo*, la que pudiera propiciar los próximos pasos adelante en nuestro camino hacia la comprensión de la naturaleza de la conciencia, sino el tal vez menos sonoro trabajo de tender puentes entre disciplinas para integrar de forma coherente una enorme cantidad de datos empíricos y propuestas teóricas. La crítica a la que sometemos en la parte segunda al separatismo representacionalista y al inseparatismo fenoménico –que pretenden idénticamente moverse en un terreno a medio camino entre lo ontológico y lo explicativo, aunque sin ofrecer finalmente nada en este segundo respecto–, ha de ser contemplada desde esta perspectiva: el proyecto de reducir a trajín representacional el aspecto fenoménico de lo mental, así como el de improvisar demostraciones de su carácter fundamental y su autonomía, no sólo no han legado nada al margen de especulaciones apriorísticas, sino que, implícitamente, han venido dando pábulo a la que denominaremos *guerra interparadigmática* y, así, a la pretensión de elaborar una *teoría cognitiva del todo*, es decir, a la idea de que resolver

el problema de la conciencia será algo que quepa alcanzar antes mediante una teoría unilateral, omniexplicativa y fundamental que mediante el esfuerzo integrador al que hacíamos referencia. Esa teoría cognitiva del todo ha venido durante décadas presentándose con los ropajes del reduccionismo representacionalista. Por nuestra parte, no argumentaremos contra el representacionalismo en tanto que opción ontológica, sino en tanto que óbice para el desarrollo de una ciencia de la conciencia. Puede que lo mental se reduzca a lo representacional, como puede que no – habríamos de saber qué pretende significar «representacional» para decidirnos por una u otra alternativa, pero nadie lo tiene muy claro, siendo así que quienes creen saberlo, parecen finalmente saber cosas muy distintas–. Lo que parece, por el contrario, evidente, es que los engranajes conceptuales de la reducción ontológica de lo mental a lo representacional en escasa medida podrán contribuir al señalado esfuerzo integrador (primeramente, por la inoperancia explicativa que comporta su orientación ontologizante, pero también por su unilateralidad y su carácter fundamentalista y reduccionista), cosa que en ningún caso implica que la incuria en ese esfuerzo respecto de marcos teóricos de raigambre cognitivista vaya a resultar provechosa.

Aludíamos poco más arriba a la enorme vastedad y pluralidad de la conciencia. La misma aparece en la bibliografía contemporánea vinculada al supuesto de que existen muchos problemas de la conciencia. Uno de los principales objetivos de esta tesis es el de matizar este vínculo. Se habla a menudo de diferentes problemas de la conciencia: el ontológico, el causal, el funcional, etc. Nos referiremos a todos ellos en la primera parte y en la segunda defenderemos que en realidad se trata de etiquetas vacías. Como insinuábamos, sólo existe un problema de la conciencia: el de explicarla científicamente. Antes de lograr tal cosa, la discusión nominal acerca de qué sea la conciencia (el así llamado problema ontológico) se presenta como un precipitado e ineficaz escamoteo anticipatorio; después de ello lo hará como una tarea superflua. Incidiremos, pues, en el sentido en que el problema ontológico se reduce al explicativo, no entrando a detallar el modo en que el resto de los que aparecen designados de diversas maneras en la bibliografía lo hacen asimismo (por considerarlo trivial: cuando no caen directamente dentro del radio de acción del flanco explicativo histórico o el funcional, lo hacen oblicuamente como trasuntos ontologizados de alguna de sus ramas). La vastedad y pluralidad del problema de la conciencia no reside en que existan muchos problemas de la conciencia (uno ontológico, otro explicativo, otro funcional, otro causal, etc.), sino en que existen muchos fenómenos que explicar, muchos problemas explicativos, dado que nada en el

reino de lo mental se encuentra interrelacionado de forma más rica y compleja con cada clase de proceso psicológico que la experiencia consciente.

Es precisamente nuestra intención de poner de relieve el carácter espurio del problema ontológico la que justifica el extenso análisis de la filosofía de la mente de John R. Searle con el que abrimos la tercera parte de esta tesis. Su ontología de la conciencia ha sido la más detallada y comentada entre las elaboradas hasta la fecha, pero comprobaremos que no consiste sino en un cúmulo de metáforas e "hipótesis" inconfutables que, por añadidura, resultan equívocas y hasta contradictorias dentro del propio aparato conceptual que configuran. Al igual que el resto de "soluciones ontológicas", ofrece intuiciones metafísicas con más o menos gancho, ejemplos, analogías y metáforas más o menos expresivas, pero siempre inconfutables, inoperacionalizables y, en resumidas cuentas, tan estériles para la interpretación de la evidencia empírica disponible como para el desarrollo de nuevos marcos teóricos en los que incardinarla. De este modo, podríamos haber escogido otro autor para hacer manifiesta la escasa feracidad de esta clase de teorización. Searle ofrece, sin embargo, evidentes ventajas. Dadas la difusión y extensión de su producción, se ha visto obligado a emplear sus metáforas y analogías en gran cantidad de contextos, lo que ha servido para hacer explícito el modo en que las usa. Debemos, en cualquier caso, justificar la desproporcionada proporción del apartado crítico de esta tesis dedicada a su ontología de la conciencia, y no nos cabe hacerlo sino señalando que uno de los más relevantes entre nuestros objetivos es el de desenmascarar la vacuidad del halo de plausibilidad en el que ha venido envolviéndose la idea de que esta clase de especulación metafísica puede contribuir en algún sentido a resolver el problema de la conciencia, y que el caso de Searle es el más oportuno a tal fin a causa de su extendida difusión, su aparente coherencia y su prolijidad. Contraponer una definición abstracta a otra dada en una discusión acerca del estatus ontológico de los gametos o las circunferencias seguirá sin conducir a ninguna parte mientras esas definiciones guarden silencio acerca de la dinámica de la meiosis o el valor de $\pi$. Nada diferente de este silencio ha ofrecido hasta ahora el conato de definir qué sea la conciencia, el de abordarla desde el flanco ontológico.

Tal y como indicábamos en el resumen, la parte tercera incluye, además de una extensa crítica de la ontología de Searle, un capítulo dedicado a la exposición y crítica del tratamiento del problema de la conciencia que realiza Dennett y otro al que realizan los neurofenomenólogos. ¿A qué obedece que hayamos distribuido de este modo la sección crítica de la tesis? Por una parte, a motivos de carácter sociológico, por así decir:

las tres opciones teóricas que criticamos son las más difundidas y comentadas. Por otra parte, las mismas sirven para cubrir sino ya el abanico completo de opciones teóricas disponibles, sí al menos una porción razonablemente representativa del mismo: de Searle nos serviremos para criticar, como indicábamos, la vía ontológica, y haremos lo propio con Dennett y los neurofenomenólogos para explorar las lagunas de las vías explicativas representacionalista y antirrepresentacionalista que gozan de mayor predicamento, grado de elaboración y sustento teórico –veremos, por otra parte, que tanto en el caso de Dennett como en el de los neurofenomenólogos, las propuestas explicativas aparecen entreveradas con las metodológicas, resultando éstas, generalmente, de aquéllas.

Podemos resumir nuestros objetivos en los siguientes puntos:

i._Ofrecer una visión de conjunto del área de los *Consciousness Studies*.
ii._Especificar cuál es el verdadero problema de la conciencia.
iii._Defender que no tenemos motivos para creer que sea irresoluble.
iv._Argumentar acerca del modo en que determinadas corrientes e inercias teóricas pueden favorecer o entorpecer el avance hacia su solución.
v._Criticar a la luz de esa argumentación los marcos teóricos más destacados entre los disponibles en la literatura filosófica.

En último término, nuestro objetivo es sólo uno: el de realizar un recorrido global por los *Consciousness Studies* en busca de las tendencias teóricas que puedan contribuir a solucionar el problema de la conciencia.

**Agradecimientos**

Cuanto pudiera decir sería en cualquier caso insuficiente para expresar lo que debo a mi familia, particularmente a mis padres Andrés y Mª Isabel y mi hermana Emilia, pues ello es, poco más o menos, todo. La suerte me ha rodeado también de unos amigos que no me cabría pagarle y, por si fuera poco, puso a Diana en mi camino. Podría vivir diez vidas dedicadas a buscarla y seguiría sin encontrar la forma de corresponder a mi hado.

José Antonio y Pedro son culpables sólo de los aciertos que puedan leerse en las páginas que siguen. Sin su apoyo las mismas no habrían sido escritas.

Se discute a quién corresponde el mérito de haber fundado la primera biblioteca. Nos remontemos a Asurbanipal, a Pisístrato o a quien quiera que nos remontemos, a él, y a sus continuadores, gracias.

# I

# PARTE PRIMERA

## Aproximación histórica y conceptual al problema de la conciencia en la filosofía y las ciencias contemporáneas

# CAPÍTULO 1

## EL PROBLEMA DE LA CONCIENCIA. ANTES, DURANTE Y DESPUÉS DE LA REVOLUCIÓN COGNITIVA

### 1._Cuál es el problema de la conciencia

No es infrecuente que el problema de la conciencia nos sea presentado en los textos de filosofía de la mente y, en general, en los de ciencias cognitivas en los términos popularizados por David Chalmers (vid., v. g., Buzsáki, 2006: 361; Heil, 2004: 9; Livingston, 2004: 9; McDermott, 2001: 20; Taylor, 1999: 10). Según el filósofo australiano, existen una serie de problemas científicos relacionados con la mente, la conciencia y el cerebro que, por complicado que resulte su abordaje en la práctica, pueden ser considerados *fáciles*. Contarían como ejemplos de los mismos el modo en que el cerebro lleva a cabo tareas tales como discriminar entre diversos estímulos, integrar información de diferentes modalidades o emplearla para dirigir la conducta motora o la lingüística. A estos problemas fáciles contrapone Chalmers el que denomina *problema duro de la conciencia*, que define como "the question of how physical processes in the brain give rise to subjective experience" (Chalmers, 2002: 92). Todos los que Chalmers denomina problemas fáciles están relacionados con el modo en que el aparato cognitivo desempeña sus diversas funciones –de la percepción y el aprendizaje a la toma de decisiones y el lenguaje–, mientras que el problema duro –supuestamente– persistiría a pesar de que llegara el día en que lográramos desentrañar el modo en que todas y cada una de las mismas son llevadas a cabo por aquél. Aun cuando esta forma de presentar el problema de la conciencia resulte un tanto tendenciosa, la misma ha gozado de gran

popularidad y puede sernos de utilidad en este ínterin para ofrecer una primera aproximación tentativa e intuitiva al modo en que hoy se plantea el problema de la conciencia, dado que dicha forma de presentarlo conduce a uno de los núcleos fundamentales del debate contemporáneo: una vez hayamos desentrañado el modo en que funciona nuestro aparato cognitivo al completo, e incluso el modo en que cada una de las funciones desempeñadas por el mismo tienen un ancoraje material en nuestro sistema nervioso, ¿habremos explicado también el motivo por el cual ese funcionamiento neurocognitivo ha de verse acompañado por experiencias conscientes? En otras palabras, y echando mano de la habitual metáfora computacional, el problema de la conciencia tiene que ver con el hecho de que el sistema nervioso de determinados animales no sólo procesa información, sino que hace al tiempo a esos animales pacientes de un modo particular de sentir o experimentar, de vivir en primera persona al menos parte ese procesamiento, un modo particular de sentir que, en los términos popularizados por Thomas Nagel, hace que haya algo que sea como ser sujeto de una forma tal de procesamiento.

«Conciencia» (del mismo modo que «consciousness») proviene de la palabra latina «conscientia», sustantivo derivado de la voz «conscientem», participio de «conscire» (que podría traducirse por *conocer conjuntamente* o *saber junto con* otro u otros). La raíz de esta forma verbal es a su vez la del término «ciencia». Bien, en trazos ciertamente gruesos, el problema de la conciencia puede entenderse como el que entraña la tarea de reunir cabalmente esta pareja de cognados: «ciencia» y «conciencia». Simplificando talvez excesivamente, podría decirse que el núcleo del actual debate se encuentra aquí: ¿cómo abordar científicamente un *objeto* de estudio caracterizado, justamente, por su *subjetividad*? ¿Cómo explicar, ateniéndonos a los modelos estándar de explicación científica, el hecho de que algo como un sistema nervioso, que se presenta al análisis externo como algo netamente objetivo, pueda dar lugar a eventos o procesos experimentados *desde dentro*? ¿Cómo encaja la subjetividad en un mundo "material" y en una concepción "materialista" del mundo?[1]

[1] Entrecomillamos "material" y "materialista" porque ambos términos resultan ciertamente equívocos. Nuestra utilización de los mismos en este contexto obedece a criterios, digamos, retóricos: los signos que integran el famoso tesoro de Turgot se cargan de contenido con el uso, y es así que «físico» y, particularmente, «fisicalista» son términos menos utilizados en el lenguaje cotidiano y, de este modo, menos cargados y efectistas, aunque resultarían más adecuados en este contexto, pues, por ejemplo, en el cuadro que la física viene pintando del universo –núcleo de lo que denominábamos "concepción 'materialista' del mundo"– la materia sería sólo uno de los colores, al que vendrían a sumarse el espacio-tiempo, las interacciones fundamentales (nuclear fuerte, nuclear débil, electromagnética y gravitatoria), sus campos (electromagnético, gravitatorio) y las energías asociadas a aquéllas o a éstos –un complejísimo cuadro cuyo vínculo con la noción de materialismo podría incluso satirizarse aludiendo al modo en que en el mismo se halla integrada la de antimateria.

No obstante, éstos son términos nuevos para un viejo problema, porque el problema de la conciencia, a pesar de haber acompañado de un modo u otro a la filosofía y las ciencias occidentales desde los albores de nuestra cultura, no ha sido formulado en los términos en los que actualmente se debate en filosofía, psicología, neurociencias y, en general, en ciencias cognitivas hasta muy recientemente.

## 2._Antes de la revolución cognitiva

La década de los setenta puede ser contemplada hoy como el momento en el que comienzan a plantearse algunas de las cuestiones que, en poco tiempo, convergerían dando forma a la aproximación contemporánea al problema de la conciencia. Esas cuestiones, planteadas inicialmente de forma aislada, acabarían en la década de los noventa por cristalizar en el cuerpo organizado de problemas conexos que integran la agenda de una inter-disciplina a la que ha venido desde entonces denominándose *Consciousness Studies*. En este sentido, esto es, por lo que a la tardía formulación del problema de la conciencia se refiere, cabe apuntar que el mismo apenas fue vislumbrado por la filosofía moderna –a pesar de que en ella quepa rastrear antecedentes de determinados planteamientos–, dado que en el ambiente intelectual de aquel periodo el propio problema de la mente resultaba indiscernible o, cuando menos, difícilmente disociable del problema de la conciencia. Así, Descartes (vid. 1641: *Meditaciones* II, *Objeciones* IV; 1644: parte I, § 9 y ss.) concibió la mente como el acto de pensar, el cual, entendía, resulta incomprensible o contradictorio en ausencia de conciencia. También Locke, perieco filosófico del racionalista francés, ofreció una perspectiva similar al afirmar que "la idea del pensamiento en ausencia de la conciencia es tan ininteligible como la idea de un cuerpo extendido sin partes" (Locke, 1690: lib. II, cap. 1, § 19). Por otra parte, resultó complicado formular el problema de la conciencia en los términos en los que actualmente se investiga y debate hasta muy recientemente dado que, desde que Kant relegara lo mental a un ámbito trascendente, inaccesible a la investigación empírica y la matematización que toda ciencia natural requiere,[2] y hasta la caída del conductismo en psicología, la noción vino sonando sospechosa y siendo sistemáticamente desatendida en las diversas áreas de las ciencias encargadas del estudio de la mente, el cerebro y la conducta (exclusión hecha, por ejemplo, de los conatos de aproximación pergeñados por la psico-

[2] Hegel prolongaría y radicalizaría la postura kantiana, ergotizando acerca de la incapacidad de la observación empírica para el estudio tanto la mente como de la propia vida.

logía estructuralista de Titchener, la introspección sistemática de Külpe, el funcionalismo de James o la psicología de la Gestalt). "Desde Kant hasta la caída del conductismo": se trata, sin lugar a dudas, de una afirmación osada y matizable. Resulta notorio, además, que las excepciones mencionadas entre paréntesis ni son pocas ni anecdóticas. No pretendemos al bosquejar los más que pertinentes –por más que, en el caso de nuestra exposición, superficiales– matices a tan vasta afirmación trazar una exhaustiva historia de la reciente andadura de la noción de «conciencia» en las ciencias y la filosofía occidental. Nada parecido. No obstante, y de cara a matizar tan abarcadora afirmación –¡un siglo y medio en menos de una decena de palabras!–, ofreceremos algunas tímidas pinceladas acerca de la historia reciente del estudio científico de la mente y el lugar que la conciencia ocupara dentro del mismo.

En Wilhelm Maximilian Wundt suele reconocerse al fundador de la psicología científica (vid., v. g., Hergenhahn & Henley, 2014: cap. 9; Sheehy, 2004; Thagard, 2007). Fue él, según la *historia oficial*, quien emprendió la tarea de incardinar a la psicología entre el resto de las ciencias experimentales.[3] Asimismo, suele otorgarse a Wundt el mérito de haber fundado el introspeccionismo (vid., v. g., el prefacio de Wilson a Wilson & Keil, 1999: xix) contra el que se alzara, como reacción pretendidamente desmixtificadora, la escuela conductista. Wundt consideró que la conciencia, la *experiencia inmediata*, en términos de su *Grundriss der Psychologie*, constituye el objeto de la psicología, un objeto que, en sólo diecisiete años, el manifiesto conductista de Watson (1913) desterraría del campo de la psicología científica junto con toda otra noción mentalista. Wundt trató de sentar las bases para una psicología entendida como ciencia experimental, aunque una ciencia experimental en cuyo marco la posibilidad de acceso y control de las variables de interés se veía limitada, digamos, a la mitad de las mismas: los estímulos a los que un sujeto experimental era sometido podían ser seleccionados y medidos con arreglo a criterios perfectamente respetables desde el punto de vista científico, pero, ¿qué trato podían tener los investigadores con las experiencias de las que ante tales estímulos informaban los sujetos experimentales? Puede que la insatisfactoria respuesta "sencillamente tenían que fiarse de lo que los sujetos les dijeran" constituyera

[3] Bien es cierto que casi veinte años antes de que Wundt fundara el primer laboratorio de psicología experimental en la Universidad de Leipzig, Gustav Theodor Fechner había publicado ya los resultados de sus rigurosas investigaciones en psicofísica –y en este sentido es presentado en ocasiones junto con Wundt (e incluso Helmholtz) como padre de la psicología experimental (vid., v. g., Levitin, 2002: xv)–, del mismo modo que es cierto también que tanto el espíritu sistemático de la obra de Wundt como su ímproba labor en la instauración de un contexto institucional y académico para el desarrollo de la psicología científica bastan para aplicarle con justedad el epíteto de padre de la psicología científica.

un importante estímulo para la empresa de radical transformación de la psicología científica emprendida por los primeros conductistas. También el hecho de que tales informes pudieran verse contaminados por las asunciones teóricas de los sujetos experimentales, casi siempre alumnos y jóvenes investigadores ligados al famoso laboratorio de la Universidad de Leipzig, contribuía a la sospecha acerca de la objetividad y la neutralidad de los resultados obtenidos en los estudios experimentales allí realizados. Wundt abrió, pues, el camino para la elaboración de un proyecto cuyo fracaso puede entenderse como uno de los decisivos resortes que impulsaran la reacción conductista: armada con una rigurosa medición de los estímulos presentados a los sujetos experimentales y un metódico control de la situación experimental, esta psicología científica pionera trató de hallar los átomos de la mente. Cada estado consciente ocasionado por las presentaciones experimentales en el laboratorio era concebido como compuesto por partes elementales integradas en una determinada estructura. La cuidada metodología y los escrupulosos informes de sujetos experimentales entrenados para llevar a cabo certeros actos introspectivos e informar con precisión de sus experiencias durante los experimentos, se asumía, debieran ser suficientes para dar con esos átomos, esto es, con las sensaciones elementales postuladas por esta escuela estructuralista. Pero, y aquí llega el fracaso al que aludíamos como uno de los estímulos decisivos para la reacción conductista, esos átomos no aparecieron por ninguna parte, y si lo hicieron, fue sólo de forma ambigua, causando disenso y evidenciando la imposibilidad de falsar importantes segmentos de los informes de experiencias subjetivas. Este punto puede ilustrarse mediante los famosos impasses a los que llegaron las discusiones entre partidarios de diversas versiones del estructuralismo de Wundt. Así, Külpe y Titchener, alumnos ambos de Wundt, desarrollaron desde esta matriz estructuralista e introspeccionista común dos escuelas discrepantes en diversos puntos. Uno de estos puntos de desencuentro consistía en que los partidarios de cada una de dichas escuelas sostenían posturas incompatibles acerca de la implicación de la imaginería mental en los actos de pensamiento, pero ni unos ni otros disponían de medios empíricos o argumentos apodícticos de los que servirse en el establecimiento de criterios en base a los cuales decidir quién se hallaba más cerca de una acendrada descripción de los fenómenos subjetivos y, de este modo, cuando un discípulo de Külpe le explicaba a otro de Titchener que podía realizar operaciones mentales sin experimentar ni el menor rastro de imaginería mental, el debate se precipitaba indefectiblemente hacia un punto muerto tan enriquecedor y abierto a posibles abordajes experimentales como el siguiente:

TITCHENERIANO: No puedes experimentar nada parecido a eso.
KÜLPERIANO: Sí que puedo.
TITCHENERIANO: Que no.
KÜLPERIANO: Que sí.
TITCHENERIANO: ¡Que no!
KÜLPERIANO: ¡Que sí!

Otro ejemplo patentiza con igual efectividad la feracidad que prometían los debates introspeccionistas: mientras el laboratorio de Külpe informaba de unas 12.000 sensaciones elementales (los referidos átomos de la mente), el de Titchener informaba de casi 45.000.

Tampoco el funcionalismo americano coetáneo del programa estructuralista al que venimos haciendo referencia ofrecía, desde el punto de vista conductista, perspectivas mucho más halagüeñas. Bien es cierto que al morigerar el énfasis estructuralista en la introspección y conceptualizar los fenómenos mentales en estrecha relación con la conducta del organismo y su adaptación al medio, el funcionalismo avanzó hacia márgenes mentalistas menos comprometidos desde el punto de vista de la operacionalización de los fenómenos mentales y, así, desde el punto de vista de los conductistas, pero éstos siempre habrían tenido a mano una réplica como ésta: para dar con la utilidad práctica de la conciencia primero hay que lograr dar con la conciencia, y ningún método cuya validez no pueda ser fácilmente puesta en entredicho conduce a la obtención de datos acerca de la experiencia consciente. El énfasis en la introspección fue, como indicábamos, atemperado por los funcionalistas americanos, que reinterpretaron además la propia noción de introspección rompiendo con el atomismo estructuralista, pero las reservas conductistas hacia la introspección y la licitud del mundo fenomenológico al que supuestamente daba acceso seguían tan en pie frente al funcionalismo de James, Angell o Dewey como frente al estructuralismo de Wundt, Titchener o Külpe. Así pues, poco después de que Angell definiera el modo en que el funcionalismo interpreta las "operaciones de la conciencia" insistiendo en la forma en que las mismas se encuentran integradas en las "condiciones reales de vida" (Angell, 1907: 85), Watson (1913) plantea la necesidad de dar un paso más: con sumar a la conducta observable del organismo las condiciones reales de vida a las que se refería Angell, pero reinterpretadas en términos de situación estimular, tenía el conductista todos los datos que creía necesitar para la

elaboración de una ciencia psicológica que no hiciera ninguna clase de alusión a la conciencia. La metodología conductista requiere, exclusivamente, *respuestas* a *estímulos*, reacciones conductuales ante cambios en el medio en que se desenvuelve el organismo, instancias ambas perfectamente observables y mesurables, no como una sensación atómica de tono o la experiencia subjetiva asociada a la emoción del miedo. Desde el punto de vista positivista de Watson y los conductistas, el modo en que tanto estructuralistas como funcionalistas no sólo daban carta de ciudadanía en sus marcos teóricos a la noción de conciencia, sino que la situaban entre sus principales focos de interés, impedía el establecimiento de la psicología como una verdadera ciencia natural al exigir al psicólogo lidiar con entidades hipotéticas e inobservables.

Ante las difícilmente concebibles posibilidades de acceder de forma rigurosa a un ámbito que permitiera resolver de forma científicamente irreprochable los debates introspeccionistas, la reacción extrema de los conductistas puede resultar comprensible. Tal reacción dio lugar a lo que Anna Estany ha denominado "la etapa de la muerte de la conciencia" (Estany, 2005: 263), que nosotros presentábamos como un destierro: el de toda noción mentalista. Entre dichas nociones, puede entenderse la de «conciencia» como la fundamental. En este sentido, Watson propondría, en lo que hoy es considerado el manifiesto de la psicología conductista, que "la eliminación de los estados de consciencia como objetos de estudio eliminará la barrera entre la psicología y el resto de las ciencias" (Watson, 1913: 177τ) y que, por tanto, "la psicología debe descartar toda referencia a la conciencia" (Ibíd.: 163τ). El influjo de este destierro conductista se dejaría notar más allá de las lindes de la psicología científica, y durante casi medio siglo permanecerían dichas nociones mentalistas en semejante ostracismo.[4]

La intención de integrar a la psicología en el marco de las ciencias naturales contribuyó al auge del conductismo en un sentido que excede las reseñadas pretensiones de objetividad que condujeran a los conductistas al rechazo del introspeccionismo, pues dicho rechazo fue también justificado como una ampliación del ámbito de estudio de la psicología: la mente animal queda excluida de la psicología si hacemos depender a ésta de la introspección. El conductismo auspició así el renacimiento de la psicología comparada (Papini, 1999: 218-219), configurándose como una rama experimental y puramente objetiva de las ciencias naturales (Watson, 1913: 158), una ciencia biológica del

[4] Excepciones durante este periodo a esta tendencia pueden hallarse, sobre todo, fuera del ámbito norteamericano –cabe destacar en este sentido la psicología del desarrollo de Vygotsky, la de Piaget y la psicología de la percepción de los gestaltistas.

comportamiento cuyas fuentes de datos residen en la conducta observable de los organismos y en los estímulos relacionados con ella, instancias ambas mesurables y públicamente accesibles de las que partir hacia una superación de las estériles disputas introspeccionistas en un contexto en el que el acuerdo intersubjetivo podría alcanzarse mediante el recurso a rigurosos resultados experimentales. No obstante, Watson y los conductistas posteriores, a pesar de presentar su programa como eminentemente metodológico, no se limitaron a sugerir recetas acerca del modo de proceder en el laboratorio, sino que incursionaron el ámbito ontológico al plantear, por ejemplo, que el pensamiento no es otra cosa que conducta verbal subvocal acompañada de una leve actividad muscular del aparato fonador, una hipótesis exitosa al punto que atravesara las fronteras de la psicología científica para penetrar, por ejemplo, las de la antropología (vid., v. g., Linton, 1936: cap. 6), mas una hipótesis, pese a su inicial difusión y predicamento, refutada cuando en 1947 Scott M. Smith, en la Facultad de Medicina de la Universidad de Utah, paralizara temporalmente toda su actividad muscular usando tubocurarina y comprobara que su capacidad para pensar no se vio afectada por la parálisis inducida por el bloqueador neuromuscular (Smith et al., 1947, citado en Eysenck, 2000: 372 y Velmans, 2009: 61). Otro ejemplo de estas incursiones conductistas en la ontología de lo mental puede hallarse en el primer párrafo de *Behaviorism*, en el que encontramos a Watson oscilando entre metodología y ontología al referirse a la conciencia:

> Behaviourism claims that "consciousness" is neither a definite nor a usable concept. The behaviourist, who has been trained always as an experimentalist, holds, further, that belief in the existence of consciousness goes back to the ancient days of superstition and magic (Watson, 1924: 2).[5]

Este referido ámbito de la ontología de lo mental sería asimismo tangencialmente abordado por el *conductismo analítico* o *lógico*, elaborado desde el análisis del lenguaje ordinario por el filósofo oxoniense Gilbert Ryle. La filosofía oxoniense de aquel momento hacía gala de una decidida onto-fobia, y la obra de Ryle no constituyó ninguna excepción, pero sus afirmaciones acerca del estatus lógico de nuestros conceptos mentalistas son difícilmente interpretables en términos completamente neutrales desde el punto vista ontológico.

[5] En la edición revisada este fragmento fue modificado, aunque su contenido sirve igualmente para mostrar el titubeo al que aludíamos: "Behaviorism claims that 'consciousness' is neither a definable nor a usable concept; that it is merely another word for the 'soul' of more ancient times" (Watson, 1924/1930: 3).

Ryle fue un gran conocedor de la tradición fenomenológica,[6] y a pesar de ello propuso en su clásico *The Concept of Mind* un análisis de los términos mentalistas destinado a exorcizar el cartesiano fantasma en la máquina (Ryle, 1949: cap. 1) mediante una interpretación de aquéllos que los presenta como referidos a disposiciones a la conducta, de forma que, según su análisis, los propios estados mentales a los que esos términos refieren habrían de ser concebidos, precisamente, como disposiciones a la conducta.[7] Mientras que este análisis puede resultar en cierta medida verosímil para gran cantidad de términos mentalistas, resulta complicado acomodar en el marco del mismo la fenomenología de, por ejemplo, la percepción. David M. Armstrong, en su *A Materialist Theory of Mind*, intentó completar este segmento del análisis conductista, pero introduciendo elementos ajenos al espíritu original de la propuesta de Ryle. Concretamente, trató de combinar las teorías de la identidad surgidas en los años cincuenta (de las que hablaremos en el capítulo cuarto), el conductismo analítico de Ryle y una suerte de funcionalismo analítico (vid. Byrne, 1994) para defender una postura según la cual los estados mentales, incluyendo la fenomenología de la percepción (Armstrong, 1968: 226), son disposiciones a la conducta, que a su vez no son más que estados del sistema nervioso central.[8] La ontologización del análisis lógico del lenguaje mentalista emprendido por el conductismo analítico alcanzó aquí su cota máxima.

[6] Sus primeras publicaciones, en la segunda mitad de los años veinte, fueron reseñas de autores de esta tradición (su segunda publicación, por ejemplo, fue una reseña de *Ser y tiempo*). Conoció personalmente a Husserl en 1929 y asistió ese mismo año a clases de Heidegger. Impartió asimismo en los comienzos de su carrera la asignatura Logical Objectivism: Bolzano, Brentano, Husserl and Meinong. La relación de Ryle con la fenomenología, no obstante, no debe ser entendida como un escarceo de juventud: publicó a lo largo de su carrera, en un periodo que abarca casi treinta años, seis ensayos dedicados enteramente a la fenomenología, e incluyó cuatro en su compilación de artículos (vid. Ryle, 1971: vol. 1, caps. 10, 11, 12 y 13), mientras que en dicha colección sólo encontramos uno dedicado a Moore, otro a Carnap y dos a Wittgenstein. Destaquemos para acabar, con Thomasson (2002), que en su concisa autobiografía (Ryle, 1970) dedica dos páginas a sus estudios de Husserl y otros fenomenólogos, mientras que sus alusiones a Wittgenstein, Carnap y el Círculo de Viena ocupan, en total, menos de una.

[7] Como es sabido, esta misma clase de conductismo analítico ha solido atribuirse también a Wittgenstein (para diversas discusiones acerca de la pertinencia de esta atribución vid., v. g., Budd, 1989; Byrne, 2009; Vohra, 1989).

[8] A pesar del carácter híbrido y aparentemente coyuntural de esta propuesta, Armstrong elaboró un rico marco en el que, en nuestra opinión, primaron elementos que acercan más su planteamiento al defendido por los teóricos de la identidad que a cualquiera de las otras posturas que le han sido atribuidas. Así, en su clásico *The Nature of the Mind*, anterior a *A Matherialist Theory of Mind*, definía ya a la conciencia como un mecanismo de autoescaneado del sistema nervioso central (Armstrong, 1966).

## 3._La revolución cognitiva

Más allá de los enredos ontológicos y epistemológicos que el conductismo analítico se mostraba incapaz de eludir (si, por ejemplo, recordar a la tía Enriqueta consiste, meramente, en una disposición a hacer una llamada telefónica, un paciente con síndrome de enclaustramiento total o pseudocoma –condición en la que ninguna conducta motora abierta es posible– no podría recordar a su tía Enriqueta ni tener ningún otro estado mental, cuando en realidad puede),[9] el conductismo metodológico pronto se vio asediado por dificultades que llevaron no al abandono de las eficaces herramientas que ofreció a la psicología,[10] pero sí al colapso de su preponderancia. Este derrocamiento no fue ocasionado por problemas conceptuales ni por las antiintuitivas asunciones de los conductistas acerca de nuestras vidas mentales, sino antes bien por la imposibilidad de llevar a término el propio programa metodológico conductista sin abandonar los lineamientos y objetivos marcados en su hoja de ruta. De acuerdo con la forma que ésta acabara adoptando, las configuraciones estimulares a que es expuesto un organismo, sumadas a los programas de condicionamiento a que haya sido sometido, debieran ser suficientes para controlar la conducta del mismo. Del mismo modo, su historial de condicionamiento sumado a la configuración estimular presente debiera ser suficiente para predecir su conducta. Pero se da el caso de que existe una buena cantidad de variables –muchas de ellas aún "ocultas", parafraseando a Einstein– que exceden los dos factores comentados (historial y situación estimular) y que impidieron que este ideal pudiera realizarse ni aun cuando la artificialidad de las condiciones experimentales rozara el grado nulo de validez ecológica. En este sentido, llegó a hacerse evidente que alguna clase de mediadores cognitivos entre estímulos y respuestas jugaban un papel central hasta en los más sencillos paradigmas experimentales conductistas. Así, por ejemplo, Brewer (1974) demostró que incluso el condicionamiento clásico de la respuesta de parpadeo en humanos requiere un muy explícito mediador cognitivo: el conocimiento consciente por parte del sujeto de la relación entre el estímulo condicionado y la res-

[9] Recientemente se han desarrollado procedimientos que permiten a investigadores y médicos comunicarse con pacientes completamente paralizados pero completamente conscientes haciendo uso de diversas técnicas de exploración de la actividad neurofisiológica e interfaces cerebro-ordenador (vid., v. g., Birbaumer et al., 2006; 2008; Schnackers et al., 2009; Soddu et al., 2009; Sorger et al., 2009).

[10] Sólo hace falta echar un vistazo, por ejemplo, al manual de psicología del aprendizaje –área de la psicología que constituyera el principal foco de atención de los conductistas– de Michael Domjan (Domjan, 2015) para cerciorarse de que, efectivamente, sigue investigándose y trabajándose fructíferamente en el marco de paradigmas experimentales gestados en el seno de la tradición conductista.

puesta incondicionada.[11] La necesidad de introducir en el esquema estímulo-respuesta variables cognitivas había sido ya planteada casi treinta años antes por Edward C. Tolman, un conductista tan poco ortodoxo que es a menudo considerado pionero de la psicología cognitiva. En uno de sus experimentos más conocidos (Tolman, Ritchie & Kalish, 1946) demostró, al inundar un laberinto, que las ratas a las que se había enseñado a salir andando del mismo mediante técnicas de condicionamiento eran capaces de hacerlo igualmente nadando, de donde debía inferirse que no habían podido aprender a salir del mismo asociando estímulos perceptivos con respuestas musculares, dado que los grupos musculares que hubieron de utilizar para salir nadando del laberinto fueron necesariamente diferentes de los que utilizaron durante el condicionamiento en el laberinto no inundado. Tolman hizo explícita así la necesidad de recurrir a variables cognitivas, encarnadas en su caso en nociones mentalistas tales como "expectativas" o "mapas cognitivos", que abrieron camino a la revolución cognitiva al presentarse como variables intervinientes imprescindibles para la elaboración de un marco teórico capaz de dar cuenta del modo en que la conducta del organismo tiene lugar en interacción con su ambiente.

La demoledora crítica de Chomsky (1959) al libro de Skinner *Verbal Behaviour* ha sido muy comentada entre los hitos que jalonan el comentado derrocamiento del conductismo metodológico. Un aspecto poco comentado de esta crítica que resulta interesante traer a colación en este punto es el modo en que Chomsky (1959: 51) desmonta la noción conductista de estímulo, como vimos, una de las dos únicas fuentes legítimas de datos desde el punto de vista conductista. Dada la enorme variedad de aspectos de un estímulo simple ante los que puede reaccionar un organismo, y particularmente un hablante, no sabemos cuáles son los "estímulos" implicados en una determinada situación estimular hasta que el organismo responde a ellos. A partir de la crítica de Chomsky la noción de estímulo deja de referir a algo "ahí afuera" para depender de estados internos del organismo, y, en el caso de los organismos que responden hablando, deja de referir a algo "ahí afuera" para pasar a depender del modo en que el hablante interpreta los "estímulos". Esta noción de interpretación requiere ya de un sujeto activo, y la psicología cognitiva se encargaría de convertir esa actividad en manipulación o *procesamiento* de *información* –dos nociones que pronto aparecerían ya gastadas por el uso.

[11] Para una revisión actual de esta cuestión, vid. Kirsch et al. (2004).

La hegemonía conductista comenzó a periclitar, pues, con la denominada revolución cognitiva, cuyo germen hallamos en las postrimerías de los cuarenta y vemos arraigar y medrar en los cincuenta (Boden, 2006; Gardner, 1985; Miller, 2003[12]). Esta revolución trajo consigo un redescubrimiento de la mente y, de algún modo, una vuelta a la conciencia.[13] El lenguaje mentalista vuelve a ser admitido como pasible de incardinación en marcos científicamente respetables y la noción mental por excelencia, la noción de conciencia, pasa a tener así una peculiar carta de ciudadanía dentro del "paradigma informacional de la mente" –tal y como José Luis Díaz se refiere en *La conciencia viviente* al habitualmente denominado paradigma cognitivo (Díaz, 2007: 517)–. Con todo, esta vuelta cognitivista a la mente, tal y como atravesara los cincuenta, los sesenta y hasta los setenta (la era pre-PDP), no habría sino esbozado algunos tenues –aunque por lo demás decisivos– lineamientos del problema de la conciencia tal y como actualmente es concebido.

Entre los motivos que podemos hallar a la raíz de esta vuelta cognitivista a la mente pueden contarse las ya aludidas falencias del conductismo, encabezadas por la necesidad de reintroducir variables internas ajenas a la pareja estímulo-respuesta. Sin embargo, diversas influencias provenientes de disciplinas ajenas a las ciencias de la mente, la conducta y el cerebro fueron también decisivas, y es que durante la década de los cuarenta y los cincuenta se produjeron importantes avances teóricos, empíricos, metodológicos y tecnológicos en distintas áreas que parecían converger al sembrar unas simientes de las que las ciencias cognitivas sacarían poco después buen partido. Pueden en este sentido mencionarse la cibernética (reivindicada actualmente por enactivistas como un tesoro apenas aprovechado por simbolistas y conexionistas), la teoría matemática de la información (Shannon, 1948), la teoría de detección de señales –basada en investigaciones sobre radares llevadas a cabo durante los cuarenta, elaborada matemáticamente en 1954 por Peterson, Birdsall y Fox y aplicada ese mismo año a la psicología por Wilson P. Tanner, David M. Green y John A. Swets–, la teoría de control –cuya historia moderna se remonta al análisis dinámico de James Clerk Maxwell del regulador

[12] Impulsor, testigo y, al cabo de medio siglo, narrador –en el artículo citado– de dicha revolución.

[13] Cabría apuntar que el paradigma cognitivista, habitualmente concebido como sucesor del conductista, supuso una peculiar vuelta a la mente. Tanto es así, y hasta tal punto puede considerarse peculiar dicho redescubrimiento científico de la mente, que ha llegado a ser caracterizado como un "conductismo mentalista" (González Labra, 2011: 39). Dicho paradigma ha venido pues lidiando con el problema de la conciencia de forma un tanto *sui generis* y ha sido constantemente puesto en duda como marco apropiado para la explicación científica de la conciencia –de hecho, es una opinión extendida que la conciencia, entendida como experiencia subjetiva, es el único aspecto de la mente que el paradigma cognitivista se muestra irremisiblemente incapaz de afrontar con éxito.

centrífugo de Watt-Boulton– o el análisis de sistemas en ingeniería; pero, de entre los desarrollos en áreas hasta entonces ajenas a las ciencias encargadas de la mente, la conducta y el cerebro que influyeron en el advenimiento de la revolución cognitiva, la construcción de los primeros ordenadores digitales en torno a la Segunda Guerra Mundial fue el de mayor relevancia.[14]

Esta necesidad de apelar a la historia externa (Lakatos, 1970) de la psicología para dar cuenta del advenimiento del paradigma cognitivo fue perfectamente captada por Ángel Rivière cuando apuntó que "para comprender los orígenes históricos de la psicología cognitiva no sólo es necesario remontarse, en el tiempo, a la vieja tradición epistemológica de la psicología, sino también salirse de la lógica interna de la historia [de la psicología], analizando los factores externos que han dado lugar al desarrollo de un modo de hacer psicología que forma parte de un proyecto científico más general, el de la ciencia cognitiva" (Rivière, 1991: 134). Desde luego, no es éste lugar para esbozar una historia de la ciencia cognitiva.[15] No obstante, unas breves pinceladas serán de utilidad de cara a enmarcar el modo en que, como decíamos, la conciencia adquirió una peculiar carta de ciudadanía con el advenimiento de la revolución cognitiva.

A pesar de que se hable habitualmente de ciencia cognitiva en singular, dada la escasa integración de las áreas que conforman esta interdisciplina, y dada también la convivencia de varios paradigmas en su seno –simbólico, conexionista, enactivo–, consideramos más apropiado hablar de ciencias cognitivas en plural.[16] Las ciencias cognitivas están integradas por diversas áreas de distintas disciplinas interesadas en un mismo núcleo temático: la cognición, es decir, la utilización de información por organismos y sistemas, según representacionalistas; el control por parte de los organismos de sus acciones en sus medios (Dawson, Dupuis & Wilson, 2010: 39), según partidarios de posturas enactivas y corporeizadas; aquello que de forma más radical distingue al reino animal del resto de la naturaleza, esto es, la capacidad de los animales para recordar el

[14] ENIAC, inicialmente programable mediante la modificación directa del hardware utilizando cables, es considerado el primer ordenador digital de propósito general de la historia. El proyecto, financiado por el Ejercito de los Estados Unidos, comenzó a llevarse a efecto en la Escuela Moore de Ingeniería Eléctrica de la Universidad de Pensilvania en julio de 1943.

[15] Quizá la más minuciosa introducción al particular sea la que puede leerse en Boden (2006). Más compendiosa pero igualmente solvente es la de Gardner (1985).

[16] En la elección y uso del término en plural o singular han escuchado algunos ecos de la polémica acerca de la unidad de la ciencia (vid. González, 2008). Sin embargo, y a pesar de que el uso plural venga imponiéndose dado su mayor realismo en tanto descriptor del contexto institucional y la convivencia de paradigmas y áreas escasamente integradas, los hay que, como Boden (2006: vol. 1, 12), escogen la forma singular con la intención de subrayar la ineludibilidad de la interdisciplinariedad en el estudio de los fenómenos mentales, pero, por así decir, con independencia de la disposición a tomar partido por Oppenheim o por Feyerabend.

pasado, escoger en el presente y planear para el futuro (Wasserman & Zentall, 2012: 2), según la, a nuestro juicio, más atinada entre las definiciones disponibles. Diversas áreas de las neurociencias, la psicología, la lingüística, la antropología, la filosofía y la inteligencia artificial participan en este proyecto de aproximación científica a la cognición. El momento en que comenzara a fraguarse la convergencia entre áreas tan dispares ha sido fechado con mucha precisión: en el 11 de septiembre de 1956. Aquél fue el día en que se celebrara la segunda jornada del *MIT Symposium on Information Theory*. En ella, Miller presentó su clásico "Magical number seven" demostrando la existencia de unos concretísimos límites de la memoria a corto plazo en torno a siete ítems (variables en complejidad al agruparse en *chunks*); Chomsky su "Three models for the description of language", formulando la hoy famosa jerarquía de gramáticas formales y demostrando que un modelo de producción del lenguaje derivado de los planteamientos de Shannon no podría aplicarse con éxito al lenguaje natural, ante lo cual emprendiera el proyecto de su gramática transformacional; y Newell y Simon su "The logic theory machine", haciendo pública la primera demostración de un teorema de los *Principa Mathematica* de Whitehead y Russell realizada por un programa de ordenador, en agosto de ese mismo año. Esto fue suficiente para que los propios Newell y Simon (1972: 4) fechasen posteriormente el nacimiento de las ciencias cognitivas de forma tan precisa. Esta datación se popularizaría unos años después al tomarla Gardner prestada de una ponencia de Miller de junio de 1979 e introducirla en su historia de la revolución cognitiva (Gardner, 1985: 28 del original; 44 de la traducción). Con todo, y aunque esta pretensión de precisión pueda sonar a chiste, lo cierto es que en ese año de 1956 tuvieron lugar una buena cantidad de acontecimientos de enorme relevancia para la configuración del marco en que posteriormente se estudiarían los fenómenos mentales. Así, por ejemplo, Jerome S. Bruner, Jacqueline J. Goodnow y George A. Austin publican ese año *A Study of Thinking*; Claude Shannon y John McCarthy editan el clásico *Automata Studies*, y Newell y Simon, junto con Marvin Minsky, McCarthy, Shannon y Nathaniel Rochester se reúnen en la *Conferencia Dartmouth*, considerada actualmente como el evento fundacional de la inteligencia artificial.

Como señalábamos, sigue caracterizando a las ciencias cognitivas la pluralidad de enfoques y la escasa integración de disciplinas, pues, a pesar de que en diversas universidades se hayan creado ya departamentos de ciencias cognitivas, la norma es que los científicos que contribuyen al desarrollo de esta empresa interdisciplinaria sean educados y desarrollen sus carreras en departamentos de sus disciplinas de origen. Nos apro-

ximaremos sumarísimamente a continuación al modo en que desembocaran en la revolución cognitiva cada una de las mismas.

*La inteligencia artificial*

En el núcleo del programa inicial de las ciencias cognitivas se hallaba la intención de desarrollar modelos computacionales de procesos mentales o funciones cognitivas, como el razonamiento, la resolución de problemas o la comprensión del lenguaje. Las raíces de este proyecto pueden rastrearse en la concepción del pensamiento como computación, cuyos orígenes, puede decirse, son los de la lógica moderna. De este modo, George Boole (Boole, 1854) propuso un vínculo entre operaciones formales realizadas sobre conjuntos y operadores lógicos tal y como son utilizados en proposiciones, un vínculo del que pretendió extraer unas leyes del pensamiento (no en el sentido normativo de la lógica clásica, sino en un sentido descriptivo). Por aquella misma época, Charles Babage y Lady Lovelace acarician la idea de diseñar un dispositivo de cómputo (su famoso *Analytic Engine*). Babage dedicó su vida al proyecto, y desde 1833 hasta su muerte, en 1871, trabajó en el diseño de una máquina que fuera más allá de las calculadoras de Pascal y Leibniz y pudiera analizar –mediante una unidad de procesamiento controlada por tarjetas perforadas– cualquier función matemática. El proyecto de Babage fue, quizá, demasiado ambicioso para su momento, de modo que la idea de una computadora de propósito general tendría que esperar al desarrollo de la teoría de la información, la cibernética y la elaboración de modelos de implementación electrónica de las operaciones booleanas entre la década de los treinta y la de los cuarenta del pasado siglo XX.[17] El camino a recorrer en la conformación del aparato teórico y tecnológico que permitió la construcción de las primeras computadoras no fue, pues, corto, con lo cual a nadie extraña que el proyecto de Babage se demorara hasta 1943, año en que se pusiera en marcha en EEUU el ya mencionado proyecto ENIAC y en Inglaterra el Ultra, concebido para descodificar los mensajes cifrados de los alemanes durante la Segunda Guerra Mundial. Estas máquinas no pretendieron emular capacidades mentales humanas, sino sólo ser útiles en la resolución de problemas definibles algorítmicamente, pero ins-

[17] Un paso crucial en esa dirección fue dado en 1937 por Claude Shanon al demostrar que el álgebra de Boole –el medio por éste desarrollado para expresar las proposiciones lógicas y las relaciones entre ellas mediante sencillos símbolos y reglas para operar con los mismos–, expresable en términos binarios, podía usarse para describir el comportamiento de circuitos de relés y conmutadores. Este paso fue decisivo dado que con él se hizo explícita la posibilidad de implementar en una máquina operaciones lógicas.

piraron a la primera generación de científicos cognitivos interesados en hacerlo.[18] También en este momento de gestación de las ciencias cognitivas previo a su nacimiento oficial en la década de los cincuenta, de hecho, en este mismo año de 1943, el neurofisiólogo Warren McCulloch y el lógico Walter Pitts sentarían las bases para el desarrollo del conexionismo y la neurociencia computacional al proponer su modelo de neurona artificial. Su impacto en ciencias cognitivas tendría que esperar, al igual que el de la neurodinámica (Rosemblatt, 1962) y los perceptrones, al auge de la modelización de redes neuronales artificiales en la década de los ochenta. Poco después del perceptrón de Rosemblatt y la neurona de McCulloch y Pitts tendría lugar un destacado hito en la historia de la inteligencia artificial con la publicación en *Mind* del artículo "Computing machinery and intelligence", de Alan Mathison Turing (Londres 1912–Cheshire 1954),[19] a la sazón director adjunto del laboratorio de computación de la Universidad de Manchester. En dicho artículo (Turing, 1950) se plantea un interrogante que en ocasiones ha sustituido al título original en las traducciones españolas: ¿Puede pensar una máquina? La inteligencia artificial simbolista clásica se basaría en una noción de cómputo inicialmente formulada por Turing (1936), una noción de cómputo con la que el propio Turing se propuso avanzar hacia la disolución de la frontera entre máquinas y mentes. Dicha noción parte de la idea de una máquina abstracta simple capaz de simular cualquier actividad que pueda descomponerse en procedimientos explícitos, es decir, en series de transformaciones algorítmicas de cadenas de símbolos discretos. Esta noción de computación simbolista resultaba completamente ajena a la analogía conexionista con el funcionamiento del cerebro, analogía cuyo germen, como sugeríamos, hallamos asimismo antes de la década de los cincuenta. La noción de cómputo simbolista gozaría inicialmente de un mayor éxito en el emergente ámbito de la inteligencia artificial, desarrollándose en aquel momento a partir del *Logic Theorist* y el lenguaje ILP1 de Newell

[18] No obstante, John von Neumann (Neumann János Lajos), de cuyo trabajo en el proyecto ENIAC surgiera la arquitectura de programa almacenado en la que vendría a basarse la de los ordenadores digitales posteriores, se interesó –particularmente entre 1944 y 1945– en la analogía entre ordenadores y cerebros contribuyendo decisivamente al auge de la teoría de autómatas y tratando de hallar el modo de imitar el funcionamiento de éstos mediante aquéllos, un intento del que debiera haber dado cuenta en las Conferencias Silliman que la Universidad de Yale le invitara a dictar en la primavera de 1956. El cáncer que acabara con su vida en febrero del año siguiente le impidió impartirlas, pero sus notas fueron publicadas póstumamente en un texto que no deja lugar a dudas respecto de la perspectiva desde la que entendía el polímata de origen húngaro que debiera estudiarse la actividad del sistema nervioso, el cual concibe como "una máquina de calcular en el sentido estricto" y prescribe que debiera analizarse utilizando los conceptos habituales en el estudio de dichas máquinas (Von Neumann, 1958: 75 del original; 108 de la traducción).

[19] Para una resoluta panorámica de la vida y la obra de Turing puede consultarse el último capítulo de *Los lógicos*, de Jesús Mosterín (Mosterín, 2007: 369-411).

y Simon y de la agenda que McCarthy y Minsky plantearan para una disciplina que ellos mismos bautizaran en aquel entonces. Desde ese momento, máquinas y programas destinados a remedar funciones mentales como la solución de problemas o la comprensión del lenguaje vienen tratando de desentrañar la *forma* de la arquitectura de la mente.[20]

Entre algunos de los hitos del desarrollo inmediatamente posterior de la inteligencia artificial cabe mencionar el LISP, creado por McCarthy y empleado inicialmente por sus alumnos y los de Minsky para escribir programas capaces, por ejemplo, de resolver problemas de álgebra; Shakey, el robot que el equipo de Charles Rose construyera en Stanford para la solución de problemas con formas geométricas que se hallaban en su entorno; el SHRDLU de Terry Winograd, que introdujera importantes innovaciones en estructuración de datos y lograra en aquel momento el mejor interfaz lingüístico, dado que el programa era un análogo virtual de Shakey capaz de comprender órdenes que debía ejecutar con los sólidos platónicos de su entorno virtual; los *marcos* con los que Minsky y los *guiones* con los que Schank contribuyeran a avanzar hacia una inteligencia artificial capaz de tratar con información menos atomizada; y, para acabar, el General Problem Solver, con el que Newell y Simon dieran un paso más allá de la solución de problemas formales comenzando a trazar un camino que la inteligencia artificial está a día de hoy esforzándose en prolongar: el camino que va de tareas abstractas y formales como demostrar teoremas o jugar al ajedrez, relativamente fáciles de abordar computacionalmente, a tareas intuitivas, concretas, contextuales y prácticas, como, sin ir más lejos y empezando por lo más sencillo para un humano y lo más difícil para una máquina, navegar por el entorno y subir, por ejemplo, una escalera sin necesidad de leer al efecto ningún manual de instrucciones (Cortázar, 1962: 11) –no traemos este ejemplo a colación de forma gratuita, sino en alusión a la famosa caída de Asimo, el robot fruto de un proyecto para la reproducción de la conducta motora humana en el que la compañía

[20] Hablamos de *forma* por cuanto el substrato teórico de la inteligencia artificial lo encontramos en una tradición filosófica y psicológica que concibe lo mental en términos de relaciones, no de substratos: cada estado mental particular es como es dada su red de relaciones con insumos sensoriales, otros estados mentales y respuestas conductuales. La de relación, la fundamental aquí, es una noción abstracta: en una cadena de fenómenos que desemboca en "tornillo adecuadamente ajustado" la relación entre el sistema que maneja el destornillador (A) y el destornillador (B) es una relación en la que A hace que B rote, con independencia de que el agente sea un ser humano, un robot o un bonobo. Asimismo, idénticos estados mentales pueden, desde este prisma funcionalista, instanciarse en diferentes substratos: un robot y un ser humano podrían compartir idénticos diagramas de relaciones funcionales y por tanto idénticos estados mentales, pues lo mental vuelve a residir, como en Aristóteles, en la forma, no en la materia. De este modo, a nadie debe extrañar que a muchos se les presente este substrato teórico funcionalista como un redivivo dualismo en el que ya no es el alma sino determinada clase de formalismo computacional "lo que puede separarse del cuerpo como la sonrisa del gato de Cheshire" (Bartra, 2006: 22).

Honda realizara una inversión millonaria... ¡para exhibirlo en su presentación ante los medios cayendo al suelo tras tropezar en el tercer peldaño de una escalera que se disponía a subir![21]

El influjo de la inteligencia artificial en los *Consciousness Studies* es tal que no resulta en absoluto extraño encontrar en la literatura contemporánea a la conciencia definida como una forma de procesamiento de información llevada a cabo por determinada clase de máquinas. Además, dentro del área de la inteligencia artificial ha surgido el campo de investigación *Machime Consciousness*, cuyo objetivo consiste en desarrollar máquinas virtuales (Sloman & Chrisley, 2003) o robots (Holland & Goodman, 2003) autoconscientes.

*Las neurociencias*

A pesar de que el término «neurociencias» no comenzara a usarse hasta la década de los sesenta como denominación para el estudio interdisciplinario del sistema nervioso (Brook & Mandik, 2005: 3), dicho estudio, sobra indicarlo, existía antes de que se acuñara y difundiera la señalada signatura. De hecho, las disciplinas a las que hoy nos referimos con la misma experimentaron un desarrollo espectacular antes de la década de los sesenta, particularmente desde mediados del siglo XIX. Aquellos investigadores pioneros de mediados del XIX legaron principalmente descripciones histológicas y anatómicas, aunque también algunas importantes intuiciones fisiológicas. Dos hitos decisivos en esta etapa de formación de las neurociencias fueron la publicación, en 1861, de *Memoire sur le cerveau de l'homme*, obra en la que Paul Broca relacionara la afasia articulatoria del paciente que ha pasado a la historia como "Tan" con una lesión en el giro frontal inferior y, en 1865, la de la obra póstuma de Otto Deiters (muerto poco antes, con sólo 29 años) *Untersuchungen über Gehirn und Rückenmark des Menschen und der Säugethiere*. Puede partirse de cada una de ellas hacia una exposición de cada una de las dos polémicas que, paralelamente, conforman el substrato de las neurociencias modernas: la polémica entre localizacionistas y holistas, eminentemente neuropsicoló-

[21] Diferentes trabajos en robótica, como los de Rodney Brooks, Randall D. Beer, Mark Tilden o los que vienen realizándose bajo el auspicio del Departamento de Defensa de Estados Unidos en la agencia DARPA (Defence Advanced Research Project Agency), han ofrecido en las dos última décadas nuevas perspectivas acerca del modo en que, recurriendo a modelos computacionales escasamente jerárquicos y muy distribuidos, puede abordarse el reto de dotar a sistemas artificiales de sofisticadas capacidades motrices, un reto fuera del alcance de la mano –si se nos permite el juego de palabras– de la más *bruta* entre las *fuerzas* computacionales con que pudiera contar la inteligencia artificial clásica, basada en el paradigma simbólico y tendente a un procesamiento serial, jerarquizado y centralizado de la información.

gica y concerniente a la fisiología a gran escala del sistema nervioso central, y la polémica entre reticularistas y neuronistas, atinente a la anatomía y la fisiología del sistema nervioso a pequeña escala, es decir, a escala citológica e histológica. Por lo que a la primera polémica toca, el texto de Broca pudo leerse en su momento como una legitimación del localizacionismo implícito en los postulados frenológicos de Franz Joseph Gall y Johann Gaspar Spurzheim, desacreditados desde la década de los cuarenta y a aquellas alturas ausentes ya en la literatura científica y académica (aunque bien es cierto que diferentes publicaciones frenológicas, de carácter más bien popular, siguieron apareciendo hasta finales de siglo, sobre todo en Estados Unidos). Según dichos postulados, el cerebro es una comunidad similar a la descrita por Giulio Tononi en el capítulo cuarto de *Phi: A Voyage from the Brain to the Soul*, una comunidad constituida por personas aisladas, encerradas entre cuatro paredes y haciendo ellas solitas su particular trabajo. Estos postulados localizacionistas presentaban al encéfalo como algo que era posible cartografiar funcionalmente: como en un mapa, el frenólogo creía que cada facultad mental se podía localizar en una región específica del encéfalo, particularmente de la corteza cerebral. Cada una de esas facultades tendría su sede en cada una de las localizaciones que los frenólogos dibujaban aleatoriamente sobre sus famosos bustos mondos: dentro de los trazos con los que circunscribían dichas localizaciones metían a un homúnculo encargado de llevar a cabo, de forma aislada y exhaustiva, cada una de las múltiples tareas mentales que atribuían al cerebro. Así, en una línea que se remonta a la especulativa cartografía supracervical de Avicena o Alberto el Grande, que improvisaran ya en la época medieval localizaciones encefálicas más menos concretas para facultades mentales más o menos definidas, Gall postuló cerca de una treintena de facultades mentales realizadas en otros tantos órganos cerebrales, una treintena que posteriormente Spurzheim y otros ampliaran hasta la cuarentena. Dividían dichos órganos y facultades en una decena de instintos (como la amatividad, cuyo órgano hacían residir en el cerebelo, o la crueldad, que ubicaban en la porción posterior del lóbulo temporal), poco más de una decena de facultades morales (como la benevolencia, cuyo órgano ubicaban bajo el extremo de la sutura frontal), unas quince facultades perceptivas (como la prosopognosis, el de la cual situaban a ambos lados de la raíz de la nariz) y dos intelectuales (la sagacidad analógica y la causal, cuyos órganos colocaban en la parte media de la frente). El ataque a los postulados frenológicos que Marie-Jean-Pierre Flourens publicara en *Examen de la phrénologie*, de 1842, logró convencer al público culto no sólo por la posición del autor (a la sazón asiento 29 de la Académie Française y secretario permanente

de la Académie des Sciences), sino también por la sólida base empírica de su argumentación, sustentada por la reimpresión ese mismo año de *Experiences sur le système nerveux*, texto de 1825 que incluía *Recherches expérimentales sur les propriétés et les fonctions du système nerveux dans les animaux vertébrés*, de 1824, y recopilaba los pioneros estudios de ablación que ponían de manifiesto una alta inespecificidad de la relación entre localización neuroanatómica de lesiones y alteraciones de facultades concretas, una inespecificidad obviamente incompatible con los postulados frenológicos. El efecto de este contundente ataque a la ortodoxia localizacionista, apoyado en evidencia empírica obtenida en estudios de lesión realizados con palomas y conejos, se vería morigerado por las conclusiones de Broca, que, tal y como apuntábamos, situó el origen de los síntomas afásicos de Tan en su lesión en el giro frontal inferior, además de por los resultados de los experimentos de estimulación eléctrica en los que, ya en la década de los setenta, Eduard Hitzig y Gustav Theodor Fritsch descubrieron la ubicación de la corteza motora, y asimismo por las descripciones del síndrome afásico que en 1874 publicara un jovencísimo Karl Wernicke (que contaba entonces veintiséis años), unas descripciones en las que la mitad posterior de la circunvolución temporal superior aparecía como un área decisiva para el correcto funcionamiento de la facultad de la comprensión lingüística. Wernicke supo ir más allá que sus antecesores y comenzó a plantear la cuestión en términos dinámicos, en buena medida ajenos a la polémica entre localizacionistas y holistas, al esbozar un modelo de la función que dicha área contribuye a realizar en el que la misma depende de la relación entre la señalada área y otras hacia y desde las cuales ella envía y recibe eferencias y aferencias. Esta polémica fundacional terminó por diluirse al ir acumulándose evidencia contra cada uno de los bandos y comprobándose que ninguna función psicológica se realiza en una región encefálica específica, como defendían los localizacionistas, pero que cada función es resultado de la actividad de un complejo de subprocesos que, en muchos casos, sí que se realizan en regiones encefálicas específicas.

Por su parte, respecto de la polémica entre reticularistas y neuronistas, en la referida obra de Deiters aparece por primera vez descrita la estructura básica de la célula nerviosa, que comprende el cuerpo celular o soma y dos tipos de prolongaciones, a las que llamó el alemán protoplásmicas (*dendritas*, desde la acuñación de Wilhelm His en 1889) y nerviosas o cilindroejes (*axones*, desde la acuñación de Rudolf Albert von Kölliker en 1896). Por otra parte, además de esta citología descriptiva básica establecida en la década de los sesenta, desde principios de los cuarenta los trabajos de Kölliker

venían ofreciendo una panorámica de la citología nerviosa que parecía dirigirse a la concepción que aún tardaría más de medio siglo en ser generalmente aceptada: la de la neurona como unidad anatómica y funcional del sistema nervioso. Esa concepción neurohistológica acorde con los principios citológicos ya aceptados en el resto de la histología tardó en cuajar dado que en 1872 Joseph Gerlach defendió la llamada teoría reticularista –que en realidad cabe remontar a las observaciones realizadas por Franz von Leydig en 1855 sobre el tejido nervioso de arácnidos–, según la cual las prolongaciones nerviosas que Deiters describiera forman una red continua a través de la cual tiene lugar la conducción del impulso nervioso, y dado que tanto Kölliker como el resto de las figuras influyentes de la histología europea de aquel momento se adhirieron a las conclusiones de Gerlach, que creía tener en sus preparaciones microscópicas la prueba de la continuidad sin hiatos entre las fibrillas terminales de las dendritas y los cuerpos celulares vecinos que éstas alcanzan. A este modelo teórico de la microanatomía nerviosa vendría a sumarse poco después Camilo Golgi, el más pertinaz entre los reticularistas. El italiano había elaborado el método de tinción cromoargéntica que le valdría en 1906 el Nobel compartido con Santiago Ramón y Cajal y, a partir de 1873, comienza a hacer públicos los resultados que obtuvo con dicho método, los cuales conducirían a la publicación, en 1886, de su *Sulla fina anatomia degli organi centrali del sistema nervoso*. La teoría de la red no tardaría, empero, en ser desafiada desde diferentes flancos. Así, en el mismo año en que la referida obra de Golgi viera la luz aparecieron publicados los resultados de las investigaciones embriológicas realizadas por Wilhelm His, quien, desde el punto de vista histogenético, defendió la independencia de las células nerviosas al constatar que al menos algunos neuroblastos surgen y migran como células individuales. Al año siguiente August Forel cotejó la obra de Golgi con datos provenientes de la anatomía patológica y la patología experimental, cotejo que le llevaría a conjeturar la independencia de las terminaciones axónicas.[22] Estos ataques a la teoría reticular no excedieron, sin embargo, las fronteras de las hipótesis basadas en observaciones y trabajos experimentales limitados a momentos puntuales del desarrollo ontogenético y ubicaciones concretas de la anatomía nerviosa, ni tuvieron una significativa repercusión y, mientras tanto, faltaban aún los fundamentos con arreglo a los cuales elaborar un modelo citológico de la estructura del sistema nervioso en virtud del cual resultara dable esta-

[22] En sus trabajos comprobó que al seccionar prolongaciones axónicas la degeneración nerviosa no se extendía, contrariamente a lo que cabría esperar partiendo del marco reticularista, más allá de los somas de las células dañadas.

blecer una base firme para el desarrollo de la neurofisiología moderna, una nueva fisiología nerviosa basada en la traducción de impulsos eléctricos a mensajes químicos y, nuevamente, a impulsos eléctricos que nadie podía aún atisbar y en la elaboración de la cual fue necesario partir de una concepción de la comunicación nerviosa según la cual la misma tiene lugar por contacto y no por anastomosis. Serían precisamente estas lagunas las que Cajal viniera a llenar con su investigación neurohistológica. Asistido por el método de tinción cromoargéntica desarrollado por Golgi describiría Cajal entre 1887 y 1903 la microorganización del sistema nervioso prácticamente al completo y enunciaría su teoría neuronal, que terminaría definitivamente con los postulados reticularistas al presentar a la neurona como "cantón fisiológico absolutamente autónomo" (Ramón y Cajal, 1888: 9), y su ley de polarización dinámica, según la cual "la transmisión del movimiento nervioso tiene lugar desde las ramas protoplásmicas hasta el cuerpo celular, y desde éste a la expansión nerviosa" (Ramón y Cajal, 1891: 70).[23]

Sobre la anatomía y la fisiología que con su investigación iluminara Cajal se erigiría el edificio completo de lo que hoy denominamos neurociencias. Cajal ofreció a los neurocientíficos del siglo XX un campo reformado y un fructífero substrato del que no tardarían en brotar los más insospechados frutos. La década de los treinta, en la que pereciera Cajal, asistió al inicio de la investigación de los mecanismos de la transmisión nerviosa: adheridos a los planteamientos de Cajal, los neurofisiólogos de la época comienzan a desentrañar el funcionamiento del potencial de acción –una investigación cuya historia se prolonga de Luigi Galvani a las ecuaciones Hodgkin-Huxley– y a identificar neurotransmisores –una investigación que de Emil du Bois-Reymond (vid. Finger, 1994: 48) llega a nuestros días–.[24] Algunos de los grandes hitos en el campo de las neurociencias habrían de esperar, no obstante, a la década de los cuarenta, a partir de la cual se producirían destacados logros en la confluencia entre neurofisiología y psicología de la mano de Donald Hebb, David Hubel, Torsten Wiesel, Vernon Mountcastle, Roger Sperry o Wilder Penfield. No pretendemos dar aquí acabada cuenta del desarrollo de la disciplina –el volumen completo de esta tesis apenas sería suficiente a tal fin–, sino sólo apuntar sumariamente a algunos desarrollos técnicos, teóricos y empíricos especialmen-

[23] Posteriormente, Edwar Gray (Gray, 1959) desmentiría la concepción cajaliana vigente acerca del carecer trasmisor-presináptico de todo axón y receptor-postsináptico de toda dendrita. Para una serie de matizaciones y excepciones a la naturaleza de la comunicación nerviosa propuesta por Cajal consúltese Shepherd (1972), y para una perspectiva actual de dichas matizaciones y excepciones Fields (2006).

[24] Para una breve panorámica en castellano del particular consúltese Hernández Guijo (2008). Dos solventes –aunque, en el caso de la segunda, ciertamente deslavazada– introducciones a la historia de la neurotransmisión química pueden hallarse en Valenstein (2005) y Donnerer & Lembeck (2006).

te significativos durante la época de gestación de las ciencias cognitivas, esto es, precisamente a partir de esta década de los cuarenta. En cuanto a las técnicas, destaquemos que ya en la década de los treinta se abrió para las neurociencias la puerta hacia un potencial universo de datos con el diseño de la máquina electroencefalográfica multicanal de Jan Friedrich Tonnies. Otra técnica de enorme repercusión se desarrollaría poco después: el registro de la actividad de una única célula (single-unit recording), una técnica que en 1959 permitiría a Hubel y Wiesel cartografiar la corteza visual del gato con electrodos de tungsteno, un trabajo por el que recibirían el premio Nobel en 1981. Cada una de estas técnicas ofrece posibilidades muy diferentes: la primera posibilita la obtención de datos acerca de la actividad de grandes grupos neuronales; la segunda, la obtención de datos acerca de la actividad de una sola neurona. Por lo que a los hallazgos empíricos y desarrollos teóricos respecta, cabe destacar en primer lugar la publicación en 1949 de *The Organization of Behavior: A Neuropsychological Theory*, obra en la que Donald O. Hebb presenta un exitoso modelo para abordar el modo en que se crean patrones de conectividad neuronal, un modelo según el cual se refuerzan las conexiones entre neuronas que "disparan" conjuntamente. Ese mismo año Giuseppe Moruzzi y Horace W. Magoun comienzan a dar con algunas de las claves del funcionamiento del sistema reticular de activación. Al año siguiente Karl Lashley desarrolla una teoría del aprendizaje y la memoria basada en sus estudios de ablación experimental. Tanto la metodología de Lashley como sus resultados traen a las mientes los de Flourens, dado que el sentido del principio de equipotencialidad que expuso en *In Search of the Engram* puede resumirse apuntando que a cualquier área cerebral le es dable participar en la formación de memorias. Lashley llegó a dicha conclusión al comprobar que los efectos de las lesiones sobre la capacidad de aprendizaje eran mayores no cuanto más específica o circunscrita a un área determinada se hallara la lesión, sino cuanto mayor fuera la cantidad de tejido dañado. Poco después de que Lashley hiciera públicas sus conclusiones, Brenda Milner y William Beecher Scoville comenzarían a hacer lo propio con sus descripciones de la amnesia anterógrada de H. M. –sobrevenida tras la succión bilateral del hipocampo y parte de la corteza entorrinal contigua en una intervención destinada a paliar una epilepsia intratable–, ofreciendo a la neurociencia cognitiva posterior la fructífera distinción entre tipos explícitos e implícitos de memoria.[25] En aquel mismo entonces, Eugene Ase-

[25] No resulta exagerado afirmar que H. M. (cuyo nombre, Henry Gustav Molaison, se difundiera tras su muerte a finales de 2008) es el paciente neuropsicológico al que más páginas e intuiciones teóricas deben las neurociencias en general y la neurociencia cognitiva de la memoria en particular, que, de hecho, le

rinsky y Nathaniel Kleitman describen el fenómeno del movimiento rápido de los ojos (REM o MOR) durante la fase de sueño posteriormente denominada paradójica (Aserinsky & Kleitman, 1953).[26] Quizá uno de los desarrollos de mayor relevancia en este periodo sea el de la feraz investigación de los mecanismos iónicos de la iniciación y propagación del potencial de acción plasmada en el modelo que Alan Lloyd Hodgkin y Andrew Huxley publicaran en 1952. Además, señalemos de pasada y para cerrar este sucinto y fragmentario repaso histórico, esta década de los cincuenta vio dar sus primeros pasos a la investigación psicofarmacológica (para una resoluta panorámica del particular vid. Torres Bares & Escarabajal Arrieta, 2005).

La confluencia de las neurociencias con el resto de las disciplinas cognitivas comenzaría a consolidarse en 1948 con la celebración del célebre simposio de Hixon *Cerebral Mechanisms in Behavior*, al que acudieran especialistas en neurofisiología, psicología y ciencias de la computación de la talla de Karl S. Lashley, Wolfgang Köhler y Warren S. McCulloch. Uno de los más destacados frutos de esta confluencia sería el desarrollo de la neurociencia computacional. Dicho desarrollo puede entenderse en términos de *equilibrio puntuado*, dado que, a pesar de que el referido modelo inicial de neurona artificial desarrollado por McCulloch y Pitts antecediera en más de una década al *MIT Symposium on Information Theory*, el auge de la modelización computacional de redes neuronales artificiales no tendría lugar hasta la década de los ochenta,[27] en buena medida a causa de la influyente crítica de los preceptrones realizada por Marvin Minsky y Seymour Papert (Minsky & Papert, 1969).[28]

Como evidencia la enorme cantidad de teorías neurobiológicas de la conciencia elaboradas en el último cuarto de siglo, las neurociencias jugaron un papel central en los

---

debe su impulso inicial –Manns & Buffalo (2013: 1031) afirman al respecto que con él comenzó el estudio serio de los sistemas de memoria–. Scoville, que tomó la decisión de intervenir sin evidencia electrofisiológica acerca de la localización del área epileptógena (Ruiz-Vargas & López-Frutos, 2014: 204), advirtió al célebre discípulo de Sherrington Wilder Graves Penfield del cariz que tras la operación adquirieron los acontecimientos, y fue éste quien contactó con la psicóloga Brenda Milner para solicitarle que se encargara de evaluar los déficits mnemónicos de H. M.

[26] Kleitman había publicado *Sleep and Wakefulness* catorce años antes del señalado descubrimiento. Considerado el primer libro de texto sobre la ciencia del sueño, sería ampliado y actualizado en la clásica edición revisada de 1963.

[27] En este punto los dos volúmenes de la biblia del procesamiento distribuido en paralelo (Rumelhart, McClelland et al., 1986; McClelland, Rumelhart et al., 1986) constituyen el referente fundamental. Puede consultarse, como guía introductoria al marco teórico y metodológico, así como a la historia de la primera década de modelización neurocomputacional a manos de la segunda generación de conexionistas, *The Computational Brain* (Churchland & Sejnowski, 1992).

[28] A pesar de que hemos aludido sólo a la modelización de redes neuronales, lo cierto es que la modelización computacional puede atender a niveles que van desde la propia neurona considerada de forma aislada a las operaciones mentales consideradas de forma abstracta, pasando, claro, por el nivel de las redes y el de la combinación de dos o más redes (vid., v. g., Sigüenza, 1993: 52).

*Consciousness Studies* desde el nacimiento de la interdisciplina. Tendremos ocasión de constatar esta centralidad de las neurociencias en los *Consciousness Studies* cuando, en el quinto capítulo de esta primera parte, nos aproximemos a las principales teorías neurobiológicas de la conciencia.

*La lingüística*

La lingüística se situó inicialmente en el núcleo del esfuerzo interdisciplinario de las ciencias cognitivas, ocupando un lugar privilegiado en el marco del mismo desde su nacimiento, aquel 11 de septiembre de 1956 en el que Chomsky leyera su influyente "Three models for the description of language". Previamente, en las primeras décadas del siglo XX, la lingüística había sufrido un cambio de orientación al pasar de ser interpretada y practicada como una especialidad académica dedicada a la reconstrucción de la historia de los lenguajes a serlo como una disciplina científica consagrada al análisis de la estructura de los mismos. El trabajo inicial de Chomsky se encuadra pues en un marco que contribuyera decisivamente a configurar su profesor Zellig Harris al ubicar la sintaxis en el núcleo de la agenda de la disciplina, dando con ello un paso más allá de las unidades de análisis estructural elementales –como fonemas y morfemas– que habían venido ocupando a estructuralistas como Edward Sapir y positivistas como Leonard Bloomfield. El trabajo de Chomsky en aquellos años se basó en la noción de transformación de Harris, entendida como posibilidad de derivar oraciones complejas hasta reducirlas a sus estructuras elementales, y desembocó en su concepción de la gramática como un sistema generativo constituido por un conjunto de normas que permiten componer, exclusivamente, todas las frases del conjunto infinito de frases bien formadas. Fue en este contexto que la reseña del libro de Skinner *Verbal Behavior* que Chomsky publicara en 1959 produjera una revolución en el campo de la lingüística y propiciara una convergencia de ésta con la psicología que había sido abierta ya seis años antes con la celebración del seminario de verano, auspiciado por el Social Science Research Council –con fondos de la Carnegie Corporation–, que diera origen a la publicación de *Psycholinguistics: A Survey of Theory and Research Problems*, texto editado por Charles E. Osgood y Thomas A. Sebeok que dota definitivamente de oficialidad a la signatura escogida para designar a esta disciplina en ciernes desde mediados de los treinta. En esta reseña, Chomsky subraya que el paradigma conductista se mostraba incapaz de dar cuenta de la creatividad de la conducta lingüística de los sujetos, dado que limitarse a la

consideración de la reproducción por parte de éstos de respuestas adquiridas no puede dar cuenta de la capacidad humana para la producción de una cantidad sin límites de nuevas sentencias gramaticalmente bien formadas. La gran influencia de Chomsky en los años de formación de las ciencias cognitivas –hasta bien entrada la década de los sesenta– se explica en buena medida por la convergencia entre psicología y lingüística que defendió y trató de fundamentar metodológica y filosóficamente al presentar a ésta como una parte de aquélla dedicada a la investigación de un dominio cognitivo específico, de una entre las diferentes facultades mentales (Chomsky, 1980: 1). No obstante, este temprano impacto de la lingüística en las ciencias cognitivas no tiene un paralelo claro en los *Consciousness Studies*. Así, si bien el lenguaje juega un papel central en diferentes teorías contemporáneas de la conciencia, en ellas lo hace desde planteamientos que exceden el marco teórico y metodológico de la lingüística –e incluso los de la psicolingüística y la neurolingüística– para ubicarse en el de la psicología teórica o filosófica (Carruthers, 1996) al tratar, habitualmente de forma un tanto especulativa, cuestiones como, por ejemplo, la emergencia evolutiva del lenguaje asociada a la emergencia evolutiva de nuevas formas de conciencia, u otras previas y más "abstractas" como la relativa a las capacidades lingüísticas entendidas como condiciones de posibilidad de la experiencia consciente.

*La antropología*

A pesar de que inicialmente nadie dudara que el estudio científico de la cultura humana formara con justicia parte del hexágono cognitivo –metáfora espacial utilizada por Gardner (1985) para ilustrar las relaciones entre las disciplinas inmersas en la empresa interdisciplinaria de las ciencias cognitivas–, el énfasis en este vértice del hexágono ha ido perdiendo fuelle al punto que hoy es habitual encontrar que en obras generales de ciencias cognitivas no se menciona a la antropología. Así, por ejemplo, los primeros capítulos de la *MIT Encyclopedia of the Cognitive Sciences* (Wilson & Keil, 1999) están dedicados individualmente a cada una de las disciplinas cognitivas, pero entre ellas no aparece ya la antropología.

El ejemplo más destacado de un estudio antropológico de gran impacto en ciencias cognitivas es el que Berlin y Kay describieran en su libro *Basic Color Terms: Their Universality and Evolution*, de 1969. El estudio comparó la amplitud del lexicon referido a colores y las capacidades discriminatorias de sujetos pertenecientes a diferentes

grupos culturales y hablantes de diversas lenguas. Desde la década de los cincuenta (vid. Whorf, 1956) había venido planteándose la idea de que el lenguaje que hablamos y comprendemos determina o afecta ostensiblemente nuestro modo de percibir y pensar. El estudio de Berlin y Kay constató, contra esta hipótesis (conocida como la hipótesis del determinismo lingüístico o hipótesis Sapir-Whorf), que hablantes de idiomas con cantidades variables de términos para colores tenían capacidades discriminatorias muy similares. No obstante, estudios interculturales posteriores, como los de Eleanor Rosch (que publicaba entonces como Eleanor R. Heider), matizaron las conclusiones de Berlin y Kay al encontrar, por ejemplo, significativas diferencias en la memoria para colores en función de la cantidad de términos para colores disponibles en el idioma de los sujetos (vid. Heider, 1971; 1972; Heider & Olivier, 1972).

Ciertamente, gran cantidad de temas estudiados en ciencias cognitivas seguirán necesitando apoyarse en estudios antropológicos. No dejamos, pues, de incidir en la justedad de la inclusión de la antropología en el hexágono cognitivo. Sin embargo, dado el papel periférico que la disciplina ha venido jugando en ciencias cognitivas y, particularmente, dado su escaso impacto en el área de los *Consciousness Studies*, no abundaremos en las aportaciones de la antropología a la revolución cognitiva.[29]

*La filosofía*

Entre los temas tradicionales de la filosofía se encuentran algunas de las cuestiones centrales de las que se encargan las ciencias cognitivas, como la naturaleza y la forma del funcionamiento de los procesos mentales. Así, la historia toda de la filosofía ofrece propuestas relevantes para el quehacer de los científicos cognitivos. Señalemos no obstante de forma concisa, y ciñéndonos al panorama contemporáneo, que la filosofía juega diversos papeles en el ámbito de las ciencias cognitivas. Cabe destacar el de la lógica en el marco de la computación; el de la historia y la filosofía de la ciencia, a menudo aplicadas a áreas particulares de las ciencias cognitivas, como la filosofía de las neurociencias (vid., v. g., Bickle et al., 2006) o la filosofía de la inteligencia artificial (vid., v. g., Boden, 1990); el de la filosofía de la mente –signatura abarcadora donde las haya–; y el de la meta-ciencia, es decir, el de la reflexión acerca de los aspectos más abstractos y fundamentales dentro del marco teórico de una disciplina dada, una labor

[29] El lector interesado puede consultar Boden (2006, cap. 8).

cuyos frutos sazonan en forma de esclarecimiento conceptual y crítica de supuestos teóricos y metodológicos.

*La psicología*

Siendo los problemas que hoy abordan las ciencias cognitivas los que han ocupado a la psicología desde sus orígenes, y pudiendo así ser ésta comprendida como centro de gravedad de las mismas,[30] hemos enfocado nuestra exposición desde el punto de vista de dicha disciplina, con lo cual no encontramos necesario incluir en este ínterin un sucinto esbozo de su confluencia con el resto de disciplinas cognitivas y su papel en el proyecto de las ciencias cognitivas como el dedicado al resto de disciplinas. En lugar de ello, continuaremos desarrollando el modo en que el viraje del conductismo al cognitivismo contribuyera a definir el planteamiento contemporáneo del problema de la conciencia.

## 4._La conciencia durante la revolución cognitiva

Las señaladas limitaciones del conductismo metodológico, así como diversas influencias provenientes de disciplinas científicas en principio ajenas al ámbito de lo mental, contribuirían, pues, a un cambio de orientación en las ciencias de la mente, el cerebro y la conducta que Donald Olding Hebb (1960), en su discurso presidencial para la *American Psychological Association*, describiera ya haciendo uso de la noción de revolución, una *revolución cognitiva* (vid. Baars, 1986; Gardner, 1985) que encontraba ya entonces prefigurada en los trabajos de Broadbent, Festinger, Miller o Pribram y que posteriormente (Hebb, 1974) reivindicaría trazando un vínculo con la biología del que los partidarios de la revolución cognitiva venían en ese momento tratando de deshacerse al interpretar las funciones mentales en términos computacionales: “si lo decisivo acerca

[30] En este sentido, la psicología es presentada como la disciplina central y unificadora de las ciencias cognitivas en el *Dictionary of Cognitive Science: Neuroscience, Psychology, Artificial Intelligence, Linguistics, and Philosophy* (Houdé, 2004). Una forma interesante de aproximarse a la centralidad de la psicología en las ciencias cognitivas la ofrece la consideración del siguiente extremo. Desde la fundación de la revista *Cognitive Science* en 1977, las áreas de especialidad de la mayoría de los autores que publicaban en la misma fueron la psicología y la inteligencia artificial, predominando esta última inicialmente. El predominio se convertiría en equilibrio dentro de la primera década de vida de la revista, repartiéndose equitativamente psicólogos y especialistas en inteligencia artificial dos tercios del espacio. No obstante, la proporción de psicólogos continuó aumentando y actualmente más de la mitad de los autores que publican en *Cognitive Science* son psicólogos (vid. Gentner, 2010).

de la mente puede capturarse en una arquitectura computacional –diría la abrumadora mayoría de revolucionarios entre los sesenta y los ochenta–, el objeto de nuestros intereses científicos está en la *forma* de esa arquitectura antes que en los *materiales* con los que ella pueda ser levantada", motivo por el cual los filósofos de la mente funcionalistas comenzarían a hablar por aquel entonces de la posibilidad de una múltiple implementación o realización (Putnam, 1967) –en (in)diferentes soportes: sistemas nerviosos o chips de silicio, tanto da– de cualquier estado o proceso mental. El núcleo del señalado cambio de orientación residiría en la sustitución de una psicología científica basada en la conceptualización de la mente en términos de estímulos y respuestas conductuales a una psicología científica basada en una conceptualización de la mente en términos de procesamiento de información.

Los primeros diagramas elaborados en un intento por describir el flujo del procesamiento de información en el sistema cognitivo humano serían publicados por Donald Broadbent a finales de la década de los cincuenta en sus clásicos estudios de psicología de la atención (Broadbent, 1958). Se hace ya con estos primeros diagramas patente el impacto que en la formulación de las primeras teorías cognitivas tendrían desarrollos tecnológicos y científicos en áreas hasta entonces ajenas a la psicología, ya que Broadbent se basa en un marco teórico exportado del análisis de sistemas y utiliza nociones tomadas de la ingeniería eléctrica (como "filtro" o "canal"). También en la formulación de estas primeras teorías cognitivas la noción de conciencia aparecería como trasfondo problemático de los modelos elaborados en el marco de la nueva psicología del procesamiento de información. El propio modelo del filtro de la atención selectiva de Broadbent distinguía entre un procesamiento consciente y otro inconsciente, pero ni estaba destinado a explicitar estas nociones ni tenía de hecho demasiado que decir acerca del significado de cada una de ellas, sino que las despachaba tangencialmente apuntando que el procesamiento inconsciente, al contrario que el consciente, es automático, rápido y "en paralelo".

Estos modelos pioneros concebían al sujeto como activo: entre estímulos y respuestas tenía lugar una codificación, almacenamiento (aprendizaje, memoria), transformación y manipulación (pensamiento, solución de problemas, planificación) de información. La pregunta a formular en este punto, por lo que al planteamiento del problema de la conciencia toca, es la referente al vínculo de la conciencia con este paisaje de manipulación interna de información. La respuesta arrojada en este momento por modelos como el de Broadbent tiene su precedente en el modo en que James (1890) esbozara los

nexos conceptuales entre sus nociones de conciencia, memoria y atención. Según James, los contenidos presentes de la conciencia se hallan en una memoria a la que denominó "primaria" y que podemos entender hoy como "a corto plazo". En lo atinente a la relación entre atención y conciencia, James plantea que los estímulos que alcanzan la conciencia lo hacen en competición con otros que caen fuera del foco de la atención y, así, del perímetro de la conciencia. La conciencia aparece vinculada con la atención de forma análoga en el modelo de Broadbent, destinado a explicar las limitaciones en el procesamiento de información llegada simultáneamente a los órganos sensoriales en tareas de escucha dicótica. Al ingresar simultáneamente en el sistema cognitivo a través de los órganos de los sentidos una cantidad de información superior a la que puede ser manejada conscientemente, debe producirse una selección, dado que sólo podemos atender conscientemente a una fracción de la información que constantemente accede al sistema nervioso central. Broadbent planteó que esta selección comienza en una etapa temprana del procesamiento de información con un filtrado sensorial pre-consciente que realiza un análisis físico y aún no semántico de la información llegada a los órganos de los sentidos, un análisis del que se partirá en la selección de la información que podrá acceder al canal de capacidad limitada postulado en su modelo. Sólo la información que alcanza este canal de capacidad limitada puede hacerse consciente y ser objeto de un análisis más rico y profundo –este vínculo entre conciencia y atención se ha prolongado en la psicología cognitiva hasta nuestros días (vid., v. g., Baars, 1991)–. Los planteamientos de James por lo que se refiere al vínculo entre conciencia y "memoria primaria" serían desarrollados a partir de este momento (vid. Waugh & Norman, 1965) del mismo modo en que lo fueron sus planteamientos acerca del vínculo entre conciencia y atención, esto es, en modelos más interesados en la atención o la memoria que en la propia conciencia. Esta línea se prolongó a lo largo de los cincuenta y los sesenta dando lugar a una concepción en la cual se contraponía el análisis consciente de información, flexible y voluntario aunque lento y limitado en capacidad, al análisis inconsciente, rígido y automático, más allá del alcance de la voluntad del sujeto, aunque rápido y con la capacidad de manejar en paralelo gran cantidad de información.[31] Con todo, como sugeríamos, a pesar de la referencia a la conciencia en estos modelos teóricos desarrollados en los ámbitos de la psicología de la memoria y la psicología de la atención, los mismos no estaban primariamente interesados en plantear teorías acerca de la conciencia, sino acerca de la

[31] "Procesos controlados/procesos automáticos" se convertiría en la nomenclatura estándar para hacer referencia a esta contraposición (vid. Posner & Snyder, 1975).

memoria y la atención, y así, aunque se vinculara atención y memoria con conciencia, poco se decía acerca de la naturaleza de dicho vínculo, de modo que no llegó a elaborarse un marco teórico que dotara de contenido a la noción de conciencia y la articulara al ponerla en relación con el resto de procesos y aspectos del sistema cognitivo humano. No obstante, ya a principios de la década de los setenta aparecen algunas referencias a la naturaleza de ese vínculo identificando a la conciencia con una forma concreta de manipulación de información sostenida por un sistema de procesamiento central de capacidad limitada (Posner & Boies, 1971; Posner & Warren, 1972) o con un ejecutivo central ocupado en la distribución eficaz de recursos computacionales entre diferentes tareas ejecutadas simultáneamente (Shallice, 1972; Bjork, 1975).

Como venimos viendo, en el mismo momento en que comienza el viraje cognitivo, al que poco más arriba nos referíamos como un cambio de orientación desde una psicología científica basada en la conceptualización de la mente en términos de estímulos y respuestas conductuales a una psicología científica basada en una concepción informacional de la mente, empiezan a aparecer modelos en psicología de la atención y psicología de la memoria que no pueden eludir la noción de conciencia. Se trata de la época en la que la psicología comienza a hacer uso de conceptos exportados de las ciencias de la computación y la teoría de la información constituyéndose como una nueva ciencia de la mente que no tardaría en adquirir estatus de oficialidad: sólo dos años después de la aparición de los primeros diagramas de flujo de la información en el sistema humano de procesamiento por parte de Broadbent, Miller y Bruner fundan en Harvard el *Center for Cognitive Studies*, y siete años después, es decir, en 1967, transcurridos apenas nueve años desde aquellos pioneros diagramas, uno de los investigadores asociados al centro, Ulrich Neisser, publica *Cognitive Psychology*. Habitualmente considerada el acta de nacimiento de la nueva psicología cognitiva, esta obra, una introducción al emergente área del procesamiento humano de información principalmente dedicada a la atención y el reconocimiento de patrones, se convertiría en un texto de referencia para sucesivas generaciones de estudiantes, aunque ya no en Harvard, dado que el *Center for Cognitive Studies* cerraría sus puertas en 1970 tras una breve andadura de apenas una década. La memoria y la atención serían, como venimos viendo, dos de los temas centrales de esta nueva psicología cognitiva. Así, por ejemplo, si el compendio/introducción de Neisser prestaba especial atención a la atención –valga el pleonasmo–, en el año que subsiguiera a su publicación vería la luz el influyente modelo multialmacén de la memoria de Richard Atkinson y Richard Shiffrin, investigadores en Stanford, que a la sazón y parale-

lamente a Harvard venía estableciéndose como uno de los primeros centros de prestigio en los que la nueva psicología del procesamiento de información comenzaba a ocupar el núcleo de la labor docente e investigadora. Este giro desde una psicología basada en estímulos y respuestas y centrada en los procesos de aprendizaje a una psicología de corte cognitivo, basada en nociones tomadas de la ingeniería y las ciencias formales y centrada en procesos internos se prolonga hasta nuestros días, mas, evidentemente, no es éste el lugar para trazar una exhaustiva historia del particular: fungen estas breves anotaciones de carácter histórico como parte de nuestra exposición en la medida en que sirven, sencillamente, para ilustrar el carácter de la vuelta a la mente que en la psicología del siglo XX tuvo lugar de la mano del paradigma cognitivo.

En cualquier caso, como señalábamos, esta vuelta cognitivista a la mente trazaría, nada más, algunos tenues aunque significativos contornos del problema de la conciencia tal y como actualmente es concebido, pues, "despite this renewed emphasis on explaining cognitive capacities such as memory, perception and language comprehension, consciousness remained a largely neglected topic" (Blanquet, 2011: 243). Que esta vuelta cognitivista a la mente no trajera consigo una explícita reformulación del problema de la conciencia puede apreciarse ya en el hecho de que la propia palabra «conciencia», tal y como han señalado Chris Frith y Geraint Rees (Frith & Rees, 2007: 15), fue eludida por los propios científicos cognitivos –aunque se tratara de una noción implícita cuando los mismos trataban de la atención selectiva o la memoria de trabajo–, y en el hecho de que el método introspectivo de las escuelas que el conductismo trató de superar no volvería con el redescubrimiento cognitivista de lo mental a ocupar ningún papel destacado: la introspección como fuente de datos es tratada con recelo tras la revolución cognitiva, y las intuiciones introspectivas de las que el científico cognitivo pudiera hacer uso en la elaboración de marcos teóricos han de ser, en cualquier caso, confrontadas con datos conductuales obtenidos mediante refinados protocolos experimentales. Así, quizá una de las mayores aportaciones a la formulación del problema de la conciencia que el redescubrimiento cognitivista de lo mental legara fuera la demostración de la existencia de esos fenómenos mentales que John Kihlstrom bautizara como *inconsciente cognitivo* (Kihlstrom, 1987), esto es, una serie de procesos psicológicos automáticos e inconscientes que tienen lugar durante la percepción, la memoria, la atención, el razonamiento y, en general, la gestión de la conducta del organismo. De ningún modo pretendemos menospreciar el legado del cognitivismo de los sesenta y los setenta al señalar que su mayor logro consistiera en la demostración de la existencia de proce-

sos mentales inconscientes, dado que una adecuada consideración del modo en que complicadas operaciones que consideraríamos mentales con pleno derecho pueden realizarse en absoluta ausencia de experiencia consciente conduce directamente a la pregunta acerca de los motivos por los cuales no todas las operaciones mentales que necesitamos realizar para relacionarnos con nuestro medio físico y social suceden en ausencia de experiencia consciente, y esta es una forma sencilla de presentar uno de los núcleos del debate contemporáneo: ¿para qué sirve la experiencia consciente cuando complicados procesos mentales pueden realizarse en su ausencia? Sea como fuere, como indicábamos, el redescubrimiento cognitivo de lo mental habría servido para esbozar sólo algunos tenues aunque significativos contornos del problema de la conciencia tal y como actualmente es concebido, y no sería así hasta las últimas décadas del pasado siglo XX que el mismo acabara por definirse y plantearse en los términos en los que actualmente es investigado y debatido.

## 5._La década de la conciencia

La década de los noventa fue solemnemente declarada "década del cerebro" el día 18 de Julio de 1990 a las 12:11 p. m. por el entonces presidente de los Estados Unidos de América George H. W. Bush.[32] Igualmente podríamos denominarla hoy "década de la conciencia". La proliferación de teorías y discusiones filosóficas y científicas sobre la conciencia que arranca en los últimos años de la década de los ochenta para desarrollarse y establecer afianzados núcleos temáticos en la de los noventa no tiene parangón en la historia intelectual de occidente. En esta década empiezan a organizarse congresos y reuniones interdisciplinarias (no en vano, la década comienza con una curiosa *First International Conference on the Study of Consciousness Within Science*, organizada en febrero de 1990 en San Francisco por el *Bhaktivedanta Institute*, un evento en el que participaran, entre otros, John R. Searle, Sir John C. Eccles y Karl H. Pribram), aparecen instituciones, programas universitarios de postgrado y grupos interdisciplinarios de investigación (como la *Association for the Scientific Study of Consciousness*, inicialmente presidida por Patrick Wilken y cofundada, entre otros, por Bernard Baars, David Chalmers, David Rosenthal y Thomas Metzinger; el *Center for Consciousness Studies*,

[32] Hay que apuntar que tras varios años de iniciativas incoadas por organismos como el *Consejo Nacional Asesor de Salud Mental* o el *Consejo Asesor del Instituto Nacional de Desórdenes Neurológicos y Accidentes Cerebrovasculares* de los Estados Unidos de América (vid., v. g., Tandon, 2000; Martín-Rodríguez et al., 2004).

fundado por David Chalmers y Stuart Hameroff en la Universidad de Arizona; o *The Consciousness Research Group*, dirigido por Antti Revonsuo en el *Centre for Cognitive Neuroscience*) y se publican por primera vez en la historia revistas especializadas que aglutinan el trabajo en las diferentes áreas de investigación de interés para el estudio interdisciplinario de la conciencia (como *Psyche*, *Consciousness and Cognition* o *Journal of Consciousness Studies*). El intercambio de ideas a nivel internacional y la rápida expansión de la investigación en el área hacían pensar en una nueva y prontamente consolidada interdisciplina cuyos dominios se extendían de la neurobiología a la neurocomputación, de la filosofía de la mente a la física de partículas, de la antropología a la vida artificial, de la psicología a la robótica… Una disciplina inabarcable, pero el intercambio entre especialistas había comenzado y a principios de la década aparecían ya meridianamente claros en el horizonte los principales puntos en la agenda de aquella evanescente y multiforme interdisciplina, ya entonces tanto en ciernes como en auge. Da cuenta de la apertura del señalado intercambio entre especialistas en las diversas áreas de esta interdisciplina que pronto sería bautizada como *Consciousness Studies*, por ejemplo, la celebración en 1994 del primer congreso *Toward a Science of Consciousness*, que viene desde entonces celebrándose bianualmente a expensas del afianzado *Center for Consciousness Studies* y que reuniera en aquella ocasión a afamados especialistas (de prestigio internacional ya entonces basado, fundamentalmente, en sus contribuciones al estudio de la conciencia) provenientes de las neurociencias, la psicología, las ciencias de la computación, la física y la filosofía (Bernard Baars, David Chalmers, Owen Flanagan, Alvin Goldman, Stuart Hameroff, Christof Koch, Benjamin Libet, Roger Penrose, Karl Pribram o el polémico Gary Schwartz fueron algunos de los participantes en aquella primera reunión). Ya entonces podía decirse lo que Jeffrey Alan Gray dijera diez años después en las primeras frases del prefacio a su último libro, *Consciousness: Creeping up on the Hard Problem*:[33] que, a diferencia de lo que venía sucediendo en décadas anteriores, la conciencia estaba ya de moda. Y, ciertamente, sigue estándolo, porque, aunque pueda resultar exagerado aseverar que "la investigación sobre la conciencia es uno de los temas prioritarios en la mayor parte de las universidades occidentales" (Moya Santoyo, 1999: 200-201), es raro el curso académico en que no aparece un nuevo Companion, Guide, Handbook o cualquier otra clase de obra de referencia editada por los servicios de publicación de las más prestigiosas universidades o

[33] Publicado tres meses después de la muerte del psicólogo británico.

editoriales internacionales. Con todo, a la vista de las numerosas perplejidades y lagunas explicativas que a causa de la conciencia encontramos en determinadas áreas de la biología y la psicología, consideramos que cabe –y hasta podría decirse que sobra– afirmar que este auge editorial e investigador no constituye una "mera moda" pues, como tendremos ocasión de comprobar a lo largo de los siguientes capítulos, el trecho inhollado del camino hacia una ciencia de la experiencia consciente es, con toda probabilidad, mayor que el ya transitado.

> *A la introducción al problema de la conciencia en clave histórica que ofrece este primer capítulo añadiremos en el siguiente una serie de acotaciones de carácter conceptual que serán de utilidad para a) definir el problema de la conciencia en los términos en que actualmente se presenta en la bibliografía, y b) preparar el terreno para la redefinición de esos términos que emprenderemos en la parte segunda.*

# CAPÍTULO 2

## LA PLURALIDAD DE LA CONCIENCIA. CERO DEFINICIONES, SIETE TIPOS Y CINCO PROBLEMAS

### 1._Cero definiciones

A las dificultades históricas que ha debido enfrentar el planteamiento contemporáneo del problema de la conciencia, unas dificultades que, sumarísimamente, comentamos en el capítulo anterior, cabría añadir una dificultad de tipo conceptual o terminológico. La definición de «conciencia» es un punto crucial del debate contemporáneo en torno a la misma y, curiosamente, el acuerdo a día de hoy alcanzado respecto de dicha definición parece restringirse a la constatación de que carecemos de la misma. No contamos, pues, con nada parecido a una definición aceptada de consuno y compartida por los especialistas involucrados en la empresa de desentrañar el problema de la conciencia. Algunos llegan en este punto al extremo de afirmar que "resulta difícil decir algo preciso sobre la conciencia" (Arrollo, 2016: 10). Puede, no obstante, que la falta de una definición no resulte necesariamente problemática. Piénsese, por ejemplo, en las investigaciones de Mendel. De las mismas acabó por educirse una noción de herencia mucho más perfilada que la implícita en las intuiciones que motivaran el famoso plan sistemático de cruzamientos de *Pisum sativum*. Las ciencias, en palabras de Patricia S. Churchland y Terrence J. Sejnowski, acaban por ofrecer definiciones más que seguirse de ellas (Churchland & Sejnowski, 1992: 446).[1] En este sentido, es una práctica habitual en neu-

[1] Patricia S. Churchland ha insistido en este extremo con posterioridad, señalando: "*theories* about certain things and *definitions* as to *what in the world count* as those things evolve together, hand in hand. Firm,

rociencias, filosofía y psicología empezar por relatar los resultados de una investigación sobre la conciencia apuntando en un primer momento a la vaguedad y la amplitud del término, matizándolo después de forma somera –en ocasiones delimitándolo mediante su contraposición a términos similares–, y no preocupándose en lo sucesivo de hilar definiciones más precisas, sino de ofrecer e interpretar resultados incardinándolos en el seno marcos teóricos particulares. Así, por ejemplo, Edelman y Tononi comienzan el primer capítulo de *A Universe of Consciousness* señalando que todo el mundo sabe lo que es la conciencia –lo que nos abandona cada noche cuando caemos dormidos, dicen en lo que puede considerarse un guiño a Searle (1990b: 635; 1992: 83 del original, 95 de la traducción; 1993a: 7; 1997a: 5 del original, 19 de la traducción; 1998: 381; 2007a: 326; 2007b: 170)–, y no se preocupan por definir de forma estricta su noción, sino que dejan que la definición surja de su exposición y vaya ganado precisión conforme ésta avanza. No obstante, existen excepciones a esta tendencia a eludir tareas, digamos, lexicógrafas, como la del neurólogo Adam Zeman, que dedica el primer capítulo al completo de *Consciousness: A User's Guide* (Zeman, 2002) a un análisis lingüístico –etimológico e histórico– de «consciousness» y allegados. Con todo, el extremo a destacar en este punto es el de la inexistencia de un verdadero acuerdo en el área de los *Consciousness Studies* acerca del significado del término en cuestión (vid., v. g., Blackmore, 2005: 7; Corballis, 2007: 571; Gross, 2012: 71; Martín-Loeches, 2008: 81; Velmans, 1996: 2; 2009: 7),[2] cosa que, como sugeríamos, no supone óbice insalvable para la empresa de explicar científicamente la conciencia, aunque, ciertamente, esta ausencia de una definición unánimemente aceptada o, al menos, ampliamente comparti-

explicit definitions become available only fairly late in the game, as the science that embeds them firms up and matures" (Churchland, 2002: 133). [Cursivas en el original].
Tal y como Wolfgang Köhler lo planteara, "si Galileo y Newton y las otras figuras pioneras de la Física moderna hubieran sentido una gran preocupación etimológica por la pureza de sus conceptos sobre la gravedad y la energía, en lugar de seguir adelante de un modo pragmáticamente ingenuo y felizmente despreocupado como lo hicieron, la Física nunca habria llegado a ser una ciencia. A lo cual puede añadirse la afirmación de que ni la Zoología, ni la Botánica, ni la Geología empezaron a ser ciencias a partir de unas definiciones correctas y adecuadas de las plantas y los animales" (citado en Nisbet, 1970: 7 del original, 14 de la traducción).

[2] Según, William Lycan, por ejemplo, habría al menos ocho significados de «conciencia» vigentes en el área de los *Consciousness Studies*: 1) Un organismo es consciente si y solo si tiene la capacidad de percibir, pensar, sentir, etc. 2) Un organismo es consciente y ejerce un control consciente sobre su cuerpo si y solo si está despierto, tiene estados mentales y gobierna sus acciones de acuerdo con ellos. 3) Un organismo es consciente *de* algo cuando se apercibe de ello, cuando ese algo es el objeto de un estado mental ocurrente. 4) Un estado o evento mental es consciente si y solo si el sujeto advierte dicho evento. 5) Un estado o evento es consciente si y solo si el sujeto puede informar del mismo. 6) Un organismo es consciente de un evento o suceso mental si y solo si tiene de él una conciencia introspectiva, una percepción a través del "sentido interno". 7) La conciencia es una experiencia subjetiva, es decir, disponible únicamente desde el punto de vista de la primera persona. 8) Se es auto-consciente si y solo si uno se advierte como siendo distinto de los demás (Lycan, 1996: 2-4).

da puede resultar problemática en algunos respectos. Así, por ejemplo, en no pocas ocasiones se han alzado voces denunciando que determinados modelos explican la conciencia definida en los términos del modelo que sea el caso, pero no la propia conciencia.

La ausencia de una definición compartida puede, además, desgranarse en una serie de problemas relacionados, dado que a ella se suma la existencia de una serie de nociones relacionadas con la de «conciencia» o similares a ella y habituales en el debate contemporáneo. Entre dichas nociones destacaremos solamente aquí la más semejante y habitual en el debate, una voz inglesa habitualmente concebida como sinónimo de «consciousness», pero de matices escurridizos respecto de los cuales reina, como enseguida comprobaremos, el desacuerdo. Se trata de «awareness», término ubicuo en el debate contemporáneo y de connotaciones potencialmente confundentes. Así, por ejemplo, Ballin (1989) ha vinculado este sinónimo de «consciousness» con la vigilancia o alerta respecto de sucesos externos mientras que Bunge & Ardila (1987) lo traducen por «percatación» y lo vinculan con la aprehensión por parte de un organismo de estímulos tanto internos como externos: un organismo se percata de un estímulo, en este sentido, cuando puede sentirlo y reaccionar ante él.[3] A los dos problemas referidos –falta de una definición compartida y presencia de vagas sinonimias– habría que sumar la proliferación de adjetivaciones. Valga la siguiente lista para que pueda el lector formarse una idea de la cantidad de cosas diferentes que dicen los especialistas en las diversas disciplinas implicadas en los *Consciousness Studies* cuando usan el término «conciencia»: access consciousness, phenomenal consciousness, transitive consciousness, intransitive consciousness, self-consciousness, background consciousness, perceptual consciousness, monitoring consciousness, reflective consciousness, creature consciousness, state consciousness, y la lista no acaba aquí, pues, sin ir más lejos, en ocasiones la adjetivación es múltiple (v. g., intransitive creature consciousness, phenomenal consciousness without state consciousness, conceptual self-consciousness, meta-representational self-consciousness) y en ocasiones implica alguna de aquellas vagas sinonimias (v. g, phenomenal self-awareness, implicit self-awareness, phenomenal self-acquaintance).[4] Pero que no cunda el pánico: el diálogo entre especialistas, insistamos, está abierto y no

[3] Ante la ausencia de un término castellano equivalente a «awareness», algunos (vid. Enríquez de Valenzuela, 2014a: 291) han recurrido a la distinción que, a diferencia de las lenguas baltoeslavas y germánicas, la nuestra establece entre los verbos ser y estar. Desde este punto de vista, «estar consciente» vendría a compartir el campo semántico de «consciousness» mientras «ser consciente» vendría a hacerlo con el de «awareness».

[4] No tenemos constancia de la utilización en lengua castellana de traducciones de algunas cuantas de estas etiquetas, motivo por el cual hemos preferido no traducir ninguna de ellas.

se ve gravemente entorpecido por los malabarismos terminológicos a los que este en apariencia abigarrado marco léxico da lugar.

Evitaremos entrar a comentar en detalle las dificultades a que pueden conducir la ausencia de una definición compartida de «conciencia», la intrusión de vagas sinonimias y la pluralidad de locuciones complejas. En lugar de ello, pasaremos a ocuparnos de un extremo estrechamente relacionado con la dificultad de ofrecer una definición de «conciencia»: la elucidación del modo en que el problema de la conciencia es de hecho un problema plural habida cuenta de la existencia de diferentes clases o tipos de conciencia.

## 2._Siete tipos

A las referidas dificultades –históricas y conceptuales o terminológicas– con las que la formulación del problema de la conciencia ha venido topando, habría que añadir que el problema de la conciencia no puede plantearse de forma unívoca o discreta porque, de hecho, se habla de diferentes tipos de conciencia.

1._En primer lugar, se habla de la conciencia como algo que un organismo o sistema puede poseer o algo de lo que puede carecer. La noción de *creature-consciousness* (Rosenthal, 1990) responde a la necesidad de distinguir esta acepción de «conciencia» de otras habituales en el debate contemporáneo. Al hablar de conciencia en este sentido aludimos, pues, a una propiedad intransitiva de un organismo o sistema en virtud de la cual cabe decir, por ejemplo, que un paciente en coma no es consciente.

2._Esta misma noción de *creature-consciousness* ha sido reelaborada y ampliada por Peter Carruthers (Carruthers, 2000: 10 y ss.), que habla de *transitive creature-consciousness* para referirse a los estados perceptivos de un organismo cuando el mismo es consciente de un determinado estado de cosas –en este sentido, decimos que el gato permanece escondido porque es consciente de la presencia del perro– y de *autoconciencia* para referirse a una clase de *creature-consciousness* que conceptualiza como una propiedad disposicional en virtud de la cual un organismo puede percibirse como un objeto diferente de otros (sentido débil de autoconciencia como forma de *transitive creature-consciousness* en la que el propio cuerpo es el objeto de la conciencia) o como una entidad con una vida mental continua e integrada (sentido fuerte de autoconciencia

como forma de *creature-consciousness* que Carruthers atribuye a los seres humanos y, con ciertas reservas, a los grandes simios).

Hasta aquí hemos hablado de la conciencia como propiedad del sujeto: se trataba de si el sujeto era o no consciente y de cómo lo era. En adelante hablaremos de propiedades de estados mentales: se tratará de si el estado mental es o no consciente y de cómo lo es.

3._Hacíamos en los puntos anteriores mención –por más que tangencialmente– de la distinción entre una clase *transitiva* y una *intransitiva* de conciencia. Como puede desprenderse de lo comentado hasta aquí, la primera se refiere a la conciencia cuando ésta es de algo, cuando hay un objeto al que ella se dirige, mientras la segunda hace referencia a la conciencia cuando no cabe hablar de ella como referida a un objeto. Un estado de ánimo como la melancolía podría caer en esta segunda categoría mientras el acto mental por el cual de pronto recordamos el nombre de un director de cine lo haría en la primera.

4._Una distinción habitual en el debate contemporáneo y, en cierta medida, análoga a la anterior la encontramos en la contraposición entre *conciencia de trasfondo* y *conciencia de estado*. La primera categoría alude a estados generales del organismo, como estar despierto, dormido, hipnotizado o en un estado elevado o bajo de arousal mientras la segunda hace referencia a estados mentales concretos individuados por sus contenidos. El sentido en que anunciábamos esta distinción como análoga a la anterior ha sido extendido hasta abarcar la noción de *creature-consciousness*, de tal modo que la conciencia de trasfondo ha sido en ocasiones concebida no sólo como intransitiva, sino asimismo como *creature-consciousness* (Chalmers, 2000).

5._La distinción crucial en el debate filosófico y científico acerca de la conciencia es sin duda la trazada por Ned Block entre *conciencia de acceso* y *conciencia fenoménica* (Block, 1995; 2001). La primera noción se refiere a estados mentales cuyos contenidos son accesibles al lenguaje y el pensamiento[5] mientras la segunda lo hace al carácter

[5] Block (1995: 231) define su noción de conciencia de acceso como sigue: un estado mental es consciente en este sentido si la representación de su contenido es a) inferencialmente promiscua, hallándose prepara-

subjetivo, cualitativo o experiencial de un estado mental, ése en virtud del cual, según la famosa expresión de Thomas Nagel (Nagel, 1974), hay para el sujeto algo que es como estar en ese estado. Otra forma de definir la conciencia fenoménica consiste en señalar que un estado es fenoménicamente consciente cuando dicho estado es *sentido* por el sujeto, cuando hay una forma distintiva de experimentarlo y una cualidad subjetiva asociada al mismo. Los estados mentales fenoménicamente conscientes son, pues, ésos de los que los filósofos de la mente dicen que poseen *qualia* (una noción en la que habremos de abundar en el siguiente capítulo). La forma más concisa de explicar qué entendemos por conciencia fenoménica consiste en indicar que ella es, sencillamente, la experiencia (Block, 1995; 2002: 206).

6._Otra noción habitual en el debate contemporáneo es la de *conciencia introspectiva*, concebida como un tipo de conciencia de acceso que funcionaría como una clase de percatación de segundo orden –*meta-awareness*– de un estado conciente dado (Hurlbert & Heavey, 2001; Jack & Shallice, 2001; Jack & Roepstorff, 2002; Schooler, 2002).

7._Encontramos, finalmente, a una buena cantidad de autores contemporáneos que, influidos por la tradición fenomenológica, hablan de una forma de conciencia que denominan *autoconciencia pre-reflexiva*. Según estos autores (vid, v. g., Wider, 1997; Williams, 1998; Zahavi, 1999), esta noción refiere a una forma primitiva de autoconciencia, una forma autorreferencial de conciencia de la propia experiencia subjetiva que no requiere de actos de reflexión, atención o introspección para darse.

### 3._Cinco problemas

El modo en que el problema de la conciencia se perfila como un problema plural debiera ser explícito ya con lo hasta el momento señalado. No obstante, no hemos destacado aún la médula de esa pluralidad. No lo demoremos: lo que en ella hallamos es, precisamente, que no hay un único problema de la conciencia, sino varios. Partiremos hacia el comentario de esta pluralidad incidiendo intuitivamente en lo que suele comprenderse como el núcleo d*el* problema de la conciencia para, a continuación, ampliar la

---

da para ser empleada como premisa del razonamiento, b) se encuentra disponible para el control racional de la acción y c) se encuentra, asimismo, a disposición del control racional de la conducta lingüística.

panorámica que la contemplación de ese núcleo ofrece mediante la inclusión en la misma de *los* diferentes problemas de la conciencia.

La filosofía de la mente contemporánea –particularmente la relacionada con los *Consciousness Studies*– está a punto de reverter saturada de experimentos mentales. Contribuyamos a ello echando mano de uno para *–nada más–* exponer con viveza a qué se denomina hoy "problema de la conciencia". Aunque mi cráneo albergara varias galaxias y mis estados conscientes –digamos, a falta de un verbo apropiado– dependieran de la cinemática galáctica, la dinámica estelar, la dispersión de las velocidades, etc. –cuya evolución en los distintos niveles de análisis haría las veces en este experimento mental de, por ejemplo, procesos de integración a gran escala de la actividad nerviosa mediante sincronización de fase de grupos neuronales ampliamente distribuidos, mapas neuronales interconectados por procesos de reentrada, potenciación a largo plazo… y así hasta llegar a la fusión de la membrana vesicular con la plasmática en el botón sináptico, la apertura y cierre de las bombas de sodio y potasio implicadas en los potenciales de acción, etc.–, una especie de expertos "neuroastrofísicos" que viviera dentro de mi cráneo y conociera con precisión cada una de las leyes que pudieran regir cada uno de los procesos de esa economía "astro-neuronal" dentro de mi cráneo, y, *ex hypothesi*, pudieran incluso determinar y predecir de forma extremadamente precisa la posición y velocidad de cada cuerpo en un intervalo de tiempo dado, así como el estado de los diversos campos dentro del mismo, ¿podrían deducir *qué* experimentaba yo en ese intervalo de tiempo? Pongámoselo más fácil. Supongamos, adicionalmente, que los neurocientíficos de mi tamaño han alcanzado un grado de conocimiento respecto de mi sistema "astro-nervioso" (que ellos llaman "nervioso" a secas) que no tiene nada en absoluto que envidiar al de los "neuroastrofísicos" que viven dentro de mi cráneo: han desentrañado cada aspecto del funcionamiento nervioso y han logrado relacionar de forma sutilísima la conducta manifiesta de mis conespecíficos con los aspectos relevantes de dicho funcionamiento. Imaginemos ahora que, casualmente, este grupo de científicos externo a mi cráneo descubre la existencia de los "neuroastrofísicos" que viven dentro de mi cráneo y logran dar con un medio para comunicarles todo lo que éstos no saben acerca de la relación de la "neuroastrofísica" que conocen perfectamente con mi conducta manifiesta, un aspecto de la "neuroastrofísica" que hasta entonces ni sospechaban –y que, claro, incluiría nuestro discurso acerca de sensaciones, sentimientos y demás pobladores del zoo fenomenológico–. Contando con semejante alud de datos y semejante refinamiento teó-

rico, ¿llegarían los "neuroastrofísicos" a una conclusión análoga a la alcanzada por Kris, Snaut y Sartorius en la novela de Stanisław Lem respecto de la vida mental de aquel extraño océano, esto es, podrían, sencillamente, atribuirme conciencia, es decir, desprovistos del "psicómetro" que Arthur C. Clarke imaginara en uno de sus primeros relatos breves (*Retreat from Earth*, de 1938), deducir *que*, efectivamente, experimento? No pretendemos que exista una respuesta correcta, ni siquiera sugerir que lo planteado en esta viñeta de ciencia ficción de pocos lances sea *concebible*. Con suscitar la intuición de que la película de la conciencia se proyecta para una sola persona con independencia de lo que las que le rodeen –¡o habiten!– sepan acerca de la cinta (primera pregunta), y la de que, incluso conociendo cada detalle de la misma y del modo en que interactuaría con elementos externos para poder ser proyectada, no podría inferirse si puede o no proyectarse y, decisivamente, ¡ni tan siquiera si se trata de la cinta de una película! (segunda pregunta), habríamos alcanzado la meta que nos proponíamos con este redivivo molino de Leibniz (1714: §17): la de ofrecer una sugestiva perspectiva de la perplejidad conceptual, teórica y metodológica a la que se denomina actualmente "problema de la conciencia".[6]

Planteábamos al comienzo del primer capítulo una serie de preguntas de las que nos servíamos para ilustrar el núcleo del problema de la conciencia. Volvamos sobre ellas y veamos qué tienen en común. ¿Cómo abordar científicamente un *objeto* de estudio caracterizado, justamente, por su *subjetividad*? ¿Cómo explicar el hecho de que algo como un sistema nervioso, que se presenta al análisis externo como algo netamente objetivo, pueda dar lugar a eventos o procesos experimentados *desde dentro*? ¿Cómo encaja la subjetividad en un mundo "material" y en una concepción "materialista" del mundo? Como a simple vista puede apreciarse, el elemento común a esta serie de interrogantes es la subjetividad. Ella, cabría decir, es la raíz del problema de la conciencia, que derivaría así del carácter subjetivo y privado de la experiencia consciente. Para hacer cabalmente comprensible esta concisa y quizá confusa frase conviene aclarar a qué refiere en ella el término «subjetivo». Cuando el mismo aparece en frases del estilo de

[6] Quizá el texto que mejor recoge esta perplejidad sea el citadísimo texto en el que Thomas Henry Huxley –abuelo de Aldous, bulldog de Darwin– compara a la conciencia con el genio de la lámpara: "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as [the appearance of Djin when Aladdin rubbed his lamp]" (Huxley, 1866: 370) –en la versión por la que citamos, una edición para el público académico americano revisada por el profesor de fisiología en la Universidad de Columbia Frederic Schiller Lee un lustro después de la muerte de Huxley en 1895, el original aparece deformado, siendo el texto que recogemos entre corchetes sustituido por "any other ultimate fact of nature".

las que acabamos de emplear siendo incluido en locuciones como «el carácter subjetivo de la experiencia consciente» no pretende significar o denotar "presencia de intereses o preferencias en el juicio" –como cuando aseveramos "que el clima de Granada (ESP) sea más agradable que el de Granada (EE. UU.) es un juicio subjetivo"–, sino más bien algo relacionado con el modo en que cada sujeto experimenta su "vida mental", algo relacionado con el hecho de que cada sujeto se encuentra respecto de su experiencia consciente en una situación radicalmente diferente de esa otra en la que se encuentra respecto de la experiencia consciente del resto: de la segunda tiene pistas y puede realizar inferencias mientras que la primera le es propia, privada, esto es, la experimenta *desde dentro*. Según Norton Nelkin, en filosofía de la mente pueden encontrarse al menos tres sentidos diferentes de «subjetividad» (Nelkin, 1996): como punto de vista (subjetividad epistémica), como origen de la voluntad (subjetividad volitiva) y como cualidad fenomenológica (subjetividad epistémica ligada a la ontología del sujeto en tanto tal). Si entendemos «subjetivo» en este tercer sentido, se nos presenta con claridad el modo en que la conciencia como fenómeno subjetivo se contrapone en el planteamiento contemporáneo del problema de la conciencia al tipo de fenómenos ontológicamente objetivos de los que han venido ocupándose las ciencias naturales. De este modo, el problema de la conciencia, tal y como hoy día es debatido, puede exponerse interinamente y en un primer acercamiento tentativo en términos disyuntivos, pues parece que, a juzgar por la perspectiva adoptada por una nada desdeñable cantidad de participantes en el actual debate, o bien aceptamos que la investigación de la conciencia ha de realizarse en términos objetivos y de tercera persona, renunciando aparentemente así a integrar en tal investigación el aspecto fundamental de la conciencia (su subjetividad, su ser experimentada *desde dentro*), o bien tratamos de mantener dentro del marco de nuestra investigación de la conciencia el carácter subjetivo de la misma, renunciando aparentemente así a la posibilidad de insertarla en un marco teórico "científicamente objetivo". En este sentido, William Lycan ha afirmado que "parece haber una infranqueable tensión entre una visión científica, materialista del mundo y el carácter subjetivo, fenoménico de la experiencia" (Lycan, 1996: 2τ).

El debate contemporáneo en torno al problema de la conciencia es habitualmente presentado en trazos tan gruesos, indefinidos y aproximativos como los ofrecidos con lo antedicho. Sin embargo, la realidad es que no hay una única tensión ni una única línea dividiendo actualmente el abanico de posibilidades y posturas dentro del debate con-

temporáneo en torno al problema de la conciencia, dado que, de hecho, no se debate un solo problema de la conciencia. La discusión gira no sólo en torno al apuntado núcleo central (la posibilidad de acomodar la subjetividad en un marco explicativo científico), sino que diversos extremos surgen aparejados al mismo. De entre ellos, quizá los decisivos –aquellos en función de los cuales resulta dable delinear en bosquejo los perfiles del actual debate presentado, por más que en términos aproximativos, los rasgos fundamentales de los polos de los distintos espectros que encontramos en el mismo– sean los siguientes:

1._*El problema descriptivo*, atinente a cuáles de las características o rasgos de la conciencia deben ser entendidos como definitorios. Atendiendo a este problema podemos distinguir los polos del fenomenalismo y el representacionalismo. Los autores que cabe incluir en el primero de dichos polos (autores como Block, Chalmers, Jackson, Loar, Nagel, Peacocke o Searle) inciden en el aspecto experiencial, cualitativo, vivencial o fenoménico de la conciencia como su rasgo distintivo defendiendo la autonomía e irreductibilidad de dicho rasgo a cualquier clase de actividad cognitiva representacional, mientras los autores que cabe incluir en el segundo (autores como Carruthers, Dennett, Dretske, Harman, Lycan, Rosenthal o Tye) insisten en el carácter representacional de toda actividad mental y parten del mismo hacia la elaboración de un marco descriptivo en el que integrar la experiencia consciente.

2._*El problema funcional*, que concierne al modo de abordar la pregunta acerca del papel desempeñado por la conciencia en la economía de relaciones con su entorno de los organismos conscientes. La cuestión es de enorme relevancia, dado que de no cumplir o servir a ninguna función biológica, la existencia de la conciencia constituiría una excepcional *rara avis in terris* desde el punto de vista biológico y se requerirían verdaderos malabarismos teóricos para justificar la presencia en el reino animal de un rasgo en apariencia tan sofisticado y que, a diferencia del resto de rasgos seleccionados y heredados a lo largo y ancho del mundo biológico, no cumpliera ninguna función ni contribuyera, por tanto, a favorecer de forma directa o indirecta la aptitud biológica de sus portadores. Aquí las posturas van del epifenomenalismo de Jackson (1982), según el cual la conciencia[7] no juega ningún papel en absoluto, a la ortodoxia funcionalista[8] que,

[7] Hay que matizar que Jackson habla en este punto de la conciencia de sensaciones corporales y percepciones.

plegándose a la elocuente noción fodoriana de epifobia (Fodor, 1989), defiende la eficacia causal de la conciencia.

3._*El problema causal.* Estrechamente emparentado con el anterior, pero vinculado con dificultades teóricas y conceptuales que, a diferencia de lo que sucedía con el mismo, exceden el marco de la ontología de lo biológico y se extienden por sobre el de la ontología en general. En pocas palabras, este problema tiene que ver con el hecho de que tanto si la conciencia puede ser tenida por causa de la conducta como si no, virtualmente, todas las teorías de la conciencia planteadas por filósofos y científicos contemporáneos se encuentran ante una difícil encrucijada teórica. Que la conciencia sea un epifenómeno carente de eficacia causal es ante todo antiintuitivo, pero, como señalábamos más arriba, también se hallaría en contradicción con la tendencia contemporánea a analizar los fenómenos mentales en términos de sus funciones y relaciones –mutuas y con la conducta–, y, decisivamente, asimismo en contradicción con los supuestos básicos de la teoría de la evolución. Por el contrario, si mi sensación consciente de picor de nariz causa mi conducta de rascarme la nariz, ¿hemos de suponer que la fisiología que subyace a esa sensación no es suficiente para causar dicha conducta? Sabemos que conductas complejas pueden realizarse en ausencia de conciencia fenoménica –valgan como ejemplo la ceguera histérica de la que hablara Freud (1910) o el modo en que la vía dorsal (Milner & Goodale, 1995)[9] puede hacer que nuestro párpado proteja nuestro ojo ante la acometida de una astilla aunque no la percibamos conscientemente–, con lo cual, cuando esas mismas conductas se realizan conscientemente, ¿hemos de entender que tienen dos causas, una neurofisiológica y otra mental? Este es el famoso problema de la sobredeterminación causal (vid. Kim, 1998: 43 y ss.; Gibb, Lowe & Ingthorsson, 2013).

4._*El problema del punto de ebullición* –análogo al denominado "explicativo" por van Gulick (Van Gulick, 2011) y emparentado con el que Seager denomina problema de la generación (vid. Seager, 1999: 18 y ss)–, concerniente al nivel en que situaríamos a la

[8] Téngase presente que en este punto dicha ortodoxia, por lo que respecta al problema de la conciencia, incluiría a autores como Searle, opuesto de hecho al planteamiento funcionalista en filosofía de la mente pero no a la concepción causalista del funcionamiento de la conciencia.

[9] No es para los propósitos de nuestro ejemplo especialmente relevante el debate en torno a la legitimidad del modo en que los autores postulan la existencia de dos sistemas funcionales bien diferenciados (la vía visual ventral y la dorsal), un debate que ha medrado a la luz de resultados experimentales que ponen de manifiesto la presencia de bucles de interacción entre ambos sistemas.

conciencia en la escala de complejidad de la organización de la materia:[10] como una característica fundamental de la realidad, tal y como propondría el panpsiquismo[11] –que de los clásicos de James (1909; 1911) o Whitehead (1929; 1933) se extiende hasta los últimos escritos de Strawson hijo–, o como una característica no fundamental llegada a la existencia a causa del comportamiento de entidades no conscientes. Esta segunda perspectiva ofrece la posibilidad de contemplar la conciencia bien como una propiedad o rasgo de todos los organismos vivos desde el primer peldaño de la escala filogenética, o desde el primer destello de motricidad o de animalidad, como sostienen diferentes versiones del continuismo (vid., Romanes, 1883; Shany, 2007; Shani, 2008; Sheets-Johnstone, 1998), o bien como un rasgo que surgiría más o menos abrupta e inopinadamente con la aparición de determinados organismos complejos, como sugiere un conjunto de perspectivas a las que cabría denominar cuvieristas o catastrofistas –ejemplos de este tipo de planteamientos pueden rastrearse en las posturas que los emergentistas británicos clásicos adoptaron en su oposición al panpsiquismo (vid., v. g., Broad, 1925) y leerse en su forma acabada en investigadores contemporáneos que parten de asunciones según las cuales el lenguaje (Jaynes, 1979; Bickerton, 1995), alguna clase de metarepresentación (como defienden Ramachandran[12] y los proponentes de las diversas higher-order thought theories of consciousness) o la "teoría de la mente" (Humphrey, 1983) son condiciones indispensables para la irrupción de la conciencia–. Una propuesta intermedia entre continuismo y cuvierismo es, por ejemplo, la de Derek Denton (vid. Denton, 2005; Denton et al., 2009), en la que la conciencia evoluciona junto con las que denomina emociones primordiales, tales como la sed, el hambre o el dolor –cabe hacer notar en este punto que las emociones primordiales de las que habla Denton no corresponden con el catálogo de emociones primarias propuesto desde la psicología de la emoción (alegría, asco, ira, miedo, sorpresa y tristeza).

5._*El problema ontológico*, en torno al cual se elaboraran las respuestas tradicionales al problema mente-cuerpo de las que trataremos brevemente en el capítulo cuarto. Este problema ontológico puede ser descrito como el de situar en nuestra concepción del mundo a la conciencia: ¿es ella algo independiente del mundo físico? De no ser así,

[10] Para una introducción al *emergente* campo de las ciencias de la complejidad, vid. Mitchel (2009). Para una aplicación de la teoría de niveles de complejidad a la conciencia humana, vid. Brown, Glazebrook & Baianu (2007: especialmente 474 y ss.).
[11] Con el panpsiquismo, de acuerdo con la analogía que hemos escogido para designar este problema, la conciencia bulliría desde aproximadamente -273,15 ºC (0 K).
[12] Vid. Blackmore (2006: 186-197).

¿qué clase de relaciones guarda con él y cómo debemos plantear esas relaciones entre fenómenos físicos y fenómenos conscientes? Como es sabido, a estas preguntas ha tratado de responderse tradicionalmente mediante teorías metafísicas monistas (idealistas, de la identidad, del doble aspecto, dualistas de propiedades, funcionalistas) y dualistas (dualismo sustancialista, paralelismo psicofísico, interaccionismo cuántico). Nos ocuparemos de ambas clases de teorías tras introducir, en el próximo capítulo, la dualidad entre lo fenoménico y lo intencional, una contraposición fundamental para entender el debate contemporáneo, pues en torno a ella ha venido articulándose el mismo.

> *A pesar de que lo expuesto en este capítulo y el anterior podría concebirse como suficiente para una exposición destinada a definir cuál es el problema de la conciencia y de qué modo se plantea actualmente, dejar de atender a la contraposición entre lo fenoménico y lo intencional supondría dejar de lado los términos que dan forma a la práctica totalidad de las polémicas en curso, pero, sobre todo, dificultaría la comprensión del trabajo argumentativo que desarrollamos en la segunda parte (particularmente en el capítulo octavo). Dedicamos, pues, el siguiente capítulo a dicha contraposición.*

# CAPÍTULO 3

## LAS DOS CARAS DE LA MENTE. *OBJETOS CON MENTE*: SISTEMAS QUE REPRESENTAN Y SIENTEN

### 1._Introducción: intencionalidad y conciencia

Cuando escribimos en este epígrafe "conciencia" ha de entenderse que hablamos de conciencia fenoménica, una noción que hemos presentado ya, pero que habremos de precisar en este capítulo introduciendo otra estrechamente relacionada con ella: la de *qualia*.[1] De la otra noción que aparece en el epígrafe, la de intencionalidad, nos ocuparemos enseguida. Comencemos por señalar que un estado intencional se caracteriza por referir o ser-acerca-de un determinado *estado de cosas*[2] y representarlo de un determinado modo. Lo mental es por tanto intencional en cuanto es-acerca-de, en cuanto representa. Existe un acuerdo general según el cual con estas dos nociones se agota el ámbito de lo mental: la mente representa (es intencional) y al menos algunos de sus procesos son sentidos, experimentados (he ahí su carácter fenoménico). Estas son las dos características definitorias de lo mental y, además, con ellas puede realizarse una descripción exhaustiva de todo fenómeno mental. Pocos autores contemporáneos mostrarían escrúpulos a la hora de suscribir asertos como el anterior. Este acuerdo general, no obstante, termina aquí, porque el disenso reina cuando se trata de precisar las relaciones entre ambos aspectos de lo mental. En este capítulo definiremos ambas nociones y en el segundo de la segunda parte criticaremos las distintas teorías acerca de las relaciones entre

[1] En relación con el vínculo entre la noción de *qualia* y el problema de la conciencia ha señalado Peter Hacker: "The alleged 'mystery of consciousness' is conceived to be the mystery of qualia" (Hacker, 2002: 158).
[2] Usamos, obviamente, de forma liberal e incidental esta noción wittgensteiniana.

ellas y acerca de las posibilidades de naturalizar la intencionalidad y la fenomenalidad de lo mental.

## 2._Intencionalidad

El término «intencionalidad», tal y como hoy es utilizado en el área de los *Consciousness Studies*, proviene del marco teórico en que Franz Brentano integrara esta noción escolástica en su *Psychologie vom empirischen Standpunkte*. En esta obra, Brentano habla de la intencionalidad en términos de inexistencia intencional, locución con la que quiso significar existir-en: la intencionalidad es en Brentano, pues, el modo de existir de los objetos *en* sus fenómenos mentales correspondientes, y para ella reserva expresiones tales como referencia a un contenido, objetividad inmanente o dirección hacia un objeto –esta idea de dirección proviene ya de la propia etimología de la palabra: el verbo latino «intendere», del que deriva la forma sustantiva «intentio», significa dirigirse a algún objetivo o meta–. Dejemos, en cualquier caso, que sea el propio Brentano quien, en el pasaje más citado de la obra a la que aludíamos, defina su noción de intencionalidad:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on. This intentional in-existence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves (Brentano, 1874: 88-89).

La definición brentaniana nos da la clave de la a menudo denominada *tesis de Brentano* (vid., v. g., Bartok, 2005: 15 y ss.; Brandl, 1996: 261 y ss.; Crane, 1995/2003: 73 y ss. de la traducción; García-Carpintero, 1996: 56; Rowlands, 2001: 198 y ss.; Villanueva, 1995: 387, y, particularmente, Quine, 1960: 221), según la cual la intencionalidad es la marca de lo mental: ella caracteriza a todos y sólo los fenómenos mentales (es decir, que ningún fenómeno no mental podría estar, como *ex hypothesi* lo está todo fenómeno mental, de este modo dirigido a un objeto), una tesis acompañada habitualmente del siguiente corolario: entendiendo que eso mental a lo que la tesis en cuestión alude se contrapone a lo físico, es evidente que, "en términos actuales, la tesis de Bren-

tano es la tesis según la cual no es posible naturalizar la intencionalidad" (Moya, 2004: 178). Sin embargo, son muchos hoy los filósofos y científicos que consideran viable naturalizar –en el ya sugerido sentido de apuntar a la posibilidad de obtener una caracterización exhaustiva del fenómeno que sea el caso en términos manejables por las ciencias naturales– el aspecto intencional de los estados mentales mientras conciben como difícil o imposible naturalizar su aspecto fenoménico. Volveremos sobre esta cuestión en el segundo capítulo de la segunda parte. De momento, antes que abordar las posibilidades de naturalizar intencionalidad y/o fenomenalidad y, asimismo, antes que tratar de las relaciones entre ambos aspectos de lo mental, nos interesa dejar claro cuáles son dichos aspectos, esto es, qué pretenden designar tales nociones. La de intencionalidad puede definirse en términos de posibilidad de tener un contenido. El contenido intencional de un estado mental es aquello a lo que dicho estado se refiere, aquello acerca de lo cual es. Así, por ejemplo, mi creencia según la cual "David sufre" es acerca de David. De este modo, un estado mental tiene un objeto o contenido intencional en virtud de esa propiedad de tales estados dada la cual éstos se dirigen a algo y en función de la cual valores de verdad o condiciones de satisfacción (Searle, 1983) pueden serles atribuidos, esto es, en virtud de su carácter intencional. En relación con su contenido intencional, pues, cabe decir de un estado mental que es erróneo, falso o inapropiado (David podría perfectamente no sufrir, crea yo lo que quiera que crea), con lo cual, el contenido intencional aparece como el portador de la normatividad de lo mental.

La intencionalidad de los estados mentales suele ilustrarse haciendo referencia a las actitudes proposicionales, instancias representacionales dotadas de contenido conceptual. Las creencias, los deseos, los temores y el resto de las actitudes proposicionales pueden caracterizarse como estados mentales cuyo rasgo común estriba en el mantenimiento de una actitud hacia una proposición, que encarnaría el señalado contenido conceptual. Así, "el bar está cerrado" es una proposición hacia cuyo contenido conceptual unos sujetos pueden tener una actitud de creencia ("Paco cree que el bar está cerrado"), otros de deseo ("Juan desea que el bar esté cerrado"), otros de temor ("Luisa teme que el bar esté cerrado"), etc. Desde esta perspectiva, un estado intencional tiene siempre un contenido, representa algo (en el ejemplo propuesto, el bar cerrado), pero además lo representa de determinado modo, es decir, desde determinada modalidad (que el bar esté cerrado es algo que puedo creer, desear, temer, saber, recordar, etc.).

De lo señalado hasta aquí cabe ya colegir que la de intencionalidad es una noción con un contenido amplio. Es decir, la literatura contemporánea no se limita a definir lo

intencional como representacional. Comienza, en efecto, señalando que con la voz «intencionalidad» nos referimos a una propiedad dada la cual un determinado estado de cosas puede, precisamente, referir a otro, ser sobre él, representarlo. Así, resulta habitual encontrar que, en una primera aproximación, se identifica "intentionality" con "aboutness" (vid., v. g, Dennett, 1983: 240; Searle, 1983: 1 del original, 17 de la traducción). Sin embargo, la cosa nunca se queda ahí y este ser-acerca-de aparece en la literatura acompañado de una serie de adláteres, un conjunto de características que integran la noción de intencionalidad vigente en el mercado de la filosofía de la mente y las ciencias cognitivas y que han aparecido ya entre líneas en lo antedicho. Listemos, no obstante, esta serie de características de las que la literatura ha venido dotando a la noción de intencionalidad haciendo previamente notar el carácter semanticista de las mismas, esto es, el modo en que asociar esta serie de características a la noción de intencionalidad equivale a llevarla más allá del mero "aboutness" y, así, a exceder el ámbito de la explicitación de una noción básica de representación o ser-acerca-de para alcanzar el de una explicitación de dicha noción según la cual «representar» significa hacer eso que el lenguaje humano hace. Dado el modo indiscriminado en que suelen usarse las nociones de lo intencional y lo representacional, no resulta baladí tener presente la posibilidad de establecer una distinción de esta clase entre un nivel básico en el cual «representar» significaría, nada más, ser-acerca-de, y un nivel semántico en el cual «representar» significa, como decíamos, hacer eso que el lenguaje humano hace.

Hemos visto ya que ser intencional significa tener la capacidad de ser-acerca-de (*aboutness*), serlo de forma errónea (*normatividad*: nada que no pueda poseer valores de verdad puede ser tenido por intencional, dado que nada que no pueda representar de forma errónea puede representar) y serlo en un sentido u otro, lo cual quiere decir que todo estado intencional está dotado de *aspectualidad* en tanto tiene un contenido y se refiere a él de un determinado modo: no se nos presenta nada íntegra y neutralmente, sino que cada vez que nos representamos algo, lo hacemos en alguna modalidad (lo deseamos, lo recordamos, lo percibimos o lo creemos) y desde alguna perspectiva (cuando veo un objeto no lo veo desde todos los ángulos a la vez del mismo modo que cuando pienso en mi tía pienso en la señora que dio a luz a mi prima, en la mujer de mi tío o en aquella agradable compañía aquel verano, pero no en todo esto a la vez). Sin embargo, a estas tres características de la intencionalidad ha venido añadiendo la literatura contemporánea otras con las cuales, ahora sí, entendemos que se da el salto desde la explicitación de una noción básica de representación a una caracterización chauvinista

de la intencionalidad según la cual ser-acerca-de significa hacer eso que el lenguaje humano hace. Entre las mismas nos encontramos en primer lugar con lo que en muchas ocasiones se presenta como un corolario de la aspectualidad: la individuación de grano fino de los aspectos bajo los cuales los objetos intencionales son representados, esto es, el hecho de que podamos pensar en Venus como Fósforo, es decir, en términos de la primera estrella visible al alba, sin pensar en Venus como Héspero, es decir, en términos de la última estrella visible durante la noche, y ello, claro, a pesar de que Fósforo y Héspero sean en realidad el mismo planeta. El modo en que esta característica habitualmente atribuida a la intencionalidad conduce a cierta suerte de chauvinismo semántico se hace palmario con ése otro en que la misma aparece en la literatura vinculada con la noción de intensionalidad u opacidad referencial. La intensionalidad es la imposibilidad de operar sustituciones *salva veritate*, la imposibilidad de sustituir en una proposición dada un término dado por un sinónimo suyo conservando el valor de verdad de la proposición, lo cual sucede cuando tal sustitución se realiza en proposiciones referidas a estados intencionales. Así, por ejemplo, sustituir Lewis Carroll por Charles Lutwidge Dodgson en la proposición "Lewis Carroll escribió el Jabberwocky" preserva su valor de verdad, lo cual no sucede necesariamente si ensayamos el mismo procedimiento sobre la proposición "Pedro cree que Lewis Carroll escribió el Jabberwocky", dado que Pedro puede saber que Lewis Carroll es el seudónimo de Charles Lutwidge Dodgson tan bien como puede desconocerlo. Este chauvinismo semántico culmina en la idea, defendida por Dennett (1983), según la cual la atribución de intencionalidad a un agente implica una atribución mínima de racionalidad: si el agente cree que A, y que A implica B, entonces deberá creer que B.

Son muchas las discusiones abiertas en filosofía de la mente y filosofía del lenguaje en torno a la intencionalidad. Aquí pretendemos sólo definir la noción y destacar sus rasgos relevantes por lo que a su implicación en los *Consciousness Studies* toca. En este sentido, la pregunta relevante a plantear es la de si, en efecto, la intencionalidad es la marca de lo mental. ¿Son, pues, intencionales todos los fenómenos mentales? Algunos autores han propuesto que buscar rasgos comunes a todos los fenómenos mentales puede ser una empresa condenada al fracaso (vid. Chomsky, 2000; Rorty, 1979), otros, dentro de la ortodoxia funcionalista, consideran que, efectivamente, todo fenómeno mental ha de ser, para contar como tal, un fenómeno representacional y por tanto intencional y, por último, aun otros entienden que existen dos rasgos característicamente mentales: la

intencionalidad y la conciencia fenoménica. En este tercer grupo encontramos, por una parte, teóricos según los cuales todos los estados mentales poseerían al tiempo ambos rasgos y, por otra, teóricos según los cuales existirían estados mentales que poseerían sólo uno de ambos.

Para aquellos que, como los defensores de las teorías representacionales, entienden que lo intencional agota lo mental, naturalizar la intencionalidad implicaría ofrecer una solución al problema de la conciencia. Para aquellos que entienden que existen dos características definitorias de lo mental y que, adicionalmente, la segunda –la conciencia fenoménica– es autónoma respecto de la primera y no puede así reducirse a ella, naturalizar la intencionalidad dejaría aún abierto el problema de la conciencia fenoménica. Si todo lo que hubiera que explicar fuera la comentada conciencia de acceso, como desearían los defensores de las teorías cognitivas o los de la teoría del espacio de trabajo global neuronal, el señalado problema no quedaría abierto, pero existe una extendida opinión según la cual el aspecto fenoménico de los estados mentales permanecería inexplicado aunque acertáramos a elaborar una irreprochable explicación naturalista de la intencionalidad de lo mental. Así, autores como Block (1996) o Peacocke (1983), defienden que el problema de la conciencia no se reduce al problema de la intencionalidad, porque lo intencional no es todo lo que hay que explicar ni puede todo lo mental explicarse exhaustivamente en términos intencionales. Según estos autores, la conciencia fenoménica, los *qualia*, no son resultado del funcionamiento representacional de la mente. Pero, sea como fuere, ¿qué se supone que son esa conciencia fenoménica y esos *qualia*?

### 3._Los qualia: de vuelta al problema duro

Hablar de los *qualia* es, desde nuestro punto de vista, una forma diferente de referirse a la conciencia fenoménica. Con todo, preguntemos, ¿a qué se refieren los usuarios de este término cuando lo emplean? Al tratar de responder a esta pregunta nos vemos obligados a admitir que la noción de *qualia* –una noción que Andy Clark ha denominado "la estrella del espectáculo de las teorías de la conciencia" (Clark, 2001: 171τ)– es utilizada en el marco del debate contemporáneo acerca del problema de la conciencia con una univocidad antes intuitiva que basada en un acuerdo explícito entre especialistas, pues parece que, a pesar de que las nociones de *qualia* que manejan los principales autores implicados en la discusión comparten un cierto *aire de familia*, ella, la noción –

en singular–, se yergue como una colección de intuiciones no demasiado definidas, como una especie de cajón de sastre en el que cada cual mete un poco más de lo mismo y, además, esta o aquella peculiaridad o matiz.[3] Trataremos a continuación de mostrar esa plural unidad intuitiva de la noción de *qualia* a través de las definiciones propuestas por diferentes usuarios del término, las cuales nos darán pie para introducir una serie de apreciaciones destinadas a poner de manifiesto tanto la señalada pluralidad de la noción de *qualia* como el sentido en que dentro de esa pluralidad puede hallarse una cierta uniformidad intuitiva.

Antes de comenzar a prestar brevemente atención a las referidas definiciones nos será de ayuda bosquejar una caracterización general de lo que habitualmente se entiende por *qualia* en el contexto de las ciencias cognitivas y la filosofía de la mente contemporánea. Las palabras, provenientes del latín, «*quale*» y «*qualia*» (plural neutro de «*quale*») son utilizadas en el señalado contexto por analogía con las formas, igualmente latinas, «quanta» y «quantum». Con ellas se pretenden designar determinadas propiedades de ciertos estados mentales: las propiedades cualitativas[4] o fenoménicas de los estados mentales conscientes, es decir, aquéllas en virtud de las cuales cabe decir que hay algo que es como tener esos estados mentales conscientes o estar en ellos, esto es, las propiedades de tales estados dadas las cuales queda determinado *cómo es* tenerlos o ser sujeto de los mismos. En la bibliografía disponible acerca de los *qualia* éstos son presentados a menudo como propiedades de los estados mentales conscientes que más nos vale definir negativamente apoyándonos en la distinción entre contenido intencional y aspecto cualitativo.

> Mientras que el contenido intencional de una experiencia particular (…) es un asunto del modo en que esa experiencia representa el mundo, su contenido cualitativo es un asunto de cómo es tenerla (García Suárez, 1995: 355).

A pesar de que no existe unanimidad al respecto, muchos autores consideran que determinados estados mentales carecen completamente de uno de ambos tipos de contenido. Así, es frecuente encontrar que se alude a sensaciones corporales, como el dolor o el orgasmo, como estados mentales en los que no habría más que contenido cualitativo,

[3] En este sentido, Nora Stigol señala que parte de la polémica en torno a los *qualia* gravita en la "falta de una noción compartida de *qualia* por los filósofos involucrados en [la] controversia" (Stigol, 2001: 32).
[4] El sustantivo latino «qualitas», raíz de los vocablos técnicos «*quale*» y «*qualia*», significa cualidad o manera de ser.

del mismo modo que es frecuente encontrar que se trata a las actitudes proposicionales como portadoras de, exclusivamente, contenido intencional. Sobre la posibilidad y la plausibilidad de tratar a determinados estados mentales como poseedores de uno o ambos de estos rasgos volveremos brevemente más abajo. Por lo pronto, centrémonos en las definiciones a las que aludíamos.

En *The MIT Encyclopedia of the Cognitive Sciences* encontramos que su entrada "*qualia*" (Wilson & Keil, 1999: 693) comienza señalando que los términos «*quale*» y «*qualia*» son habitualmente utilizados para caracterizar las propiedades cualitativas y experienciales de los estados mentales, y se hace eco a continuación de las divergencias entre los filósofos que entienden tales propiedades como características esenciales de todo estado mental consciente y los que tienden a atribuir tales propiedades sólo a las sensaciones o las percepciones. Tanto estas divergencias como aquellos calificativos («cualitativo» y «experiencial», a los que se suman otras expresiones de uso común dentro del debate, tales como «carácter fenomenológico»)[5] protagonizan la constante lid que en torno a los *qualia* tiene actualmente lugar. Un breve recorrido por algunos lugares destacados a lo largo de la historia de la conceptualización de la noción de *qualia* nos será sin duda de ayuda de cara a comprender el modo en que tales términos y divergencias se hallan insertos en las actuales discusiones en torno a los *qualia* y, principalmente, de cara a aclarar el sentido en que habitualmente se habla en ellas de los mismos.

La primera caracterización contemporánea[6] de la noción de *qualia* proviene de la etapa de su carrera en la que el pragmatista americano Clarence Irving Lewis se dedicó a la epistemología y, en concreto, de su libro de 1929 *Mind and the World Order. Outline of a Theory of Knowledge*. El texto merece ser citado por extenso:

> There are recognizable qualitative characters of the given, which may be repeated in different experiences, and are thus a sort of universals; I call these "qualia". But although such qualia are universals, in the sense of being recognized from one to another experience, they must be distinguished from the properties of objects. Confusion of these two is characteristic of many historical conceptions, as well as of current essence-

[5] Vid., v. g., la entrada "*qualia*" en el *Diccionario Oxford de la mente* (Gregory, 1987: 983 de la traducción), en la cual se asocia el *carácter fenomenológico* al modo en que las cosas *aparecen* al sujeto consciente.

[6] Como Crane (2000) apunta, los pragmatistas americanos usaron ya durante la segunda mitad del siglo XIX la noción para discutir acerca de las sensaciones, pero, como asimismo hace notar el autor, usaron las nociones de *quale* y *qualia* dotándolas de un significado más vago y general que el que actualmente se atribuye a dichos términos.

> theories. The quale is directly intuited, given, and is not the subject of any possible error because it is purely subjective. The property of an object is objective; the ascription of it is a judgment which may be mistaken; and what the predication of it asserts is something which transcends what could be given in any single experience (Lewis 1929: 121).

De entre los rasgos que Lewis atribuye a los *qualia* en este texto, los más destacados, y los que mantienen mayor vigencia, son los siguientes: los *qualia* constituyen características cualitativas y subjetivas de la experiencia directamente intuidas. La incorregibilidad a la que alude, así como la posibilidad de comparar *qualia* certeramente de experiencia a experiencia (universalidad) han sido puestas en duda y siguen siendo discutidas. No obstante, la noción de *qualia* introducida por Lewis en este texto puede ser contemplada hoy como la raíz del uso que actualmente se hace de la misma en los *Consciousness Studies*.

Con todo, a pesar de los puntos en común que la temprana aproximación de Lewis guarda con la noción contemporánea de *qualia*, la referencia clásica en el marco de las actuales discusiones en torno a los *qualia* es el artículo de 1974 "What is it like to be a bat?", de Thomas Nagel. En él defiende Nagel la existencia de una cualidad de los estados mentales que los métodos objetivos de la ciencia no pueden alcanzar, pues, argumenta, por lejos que éstos puedan llevarnos, parece que con ellos alcanzamos siempre y sólo un punto de vista *desde ninguna parte* (Nagel, 1986), pero nunca el punto de vista subjetivo que Nagel (1974) define como dotado de un carácter cualitativo en virtud del cual cabe decir que hay algo que es como ser sujeto de experiencias conscientes.

Los *qualia*, como hemos indicado, suelen asociarse al aspecto fenoménico de los estados mentales conscientes.[7] Trataremos ahora de hacer explícito brevemente el sentido en que este vínculo ha sido trazado en la filosofía de la mente contemporánea y, al tiempo, el sentido en el que la propuesta de Nagel según la cual hay algo que es como estar en un determinado estado mental fenoménicamente consciente puede considerarse un hito en la historia del problema de la conciencia y, asimismo, el punto de referencia clásico en el debate actual acerca de los *qualia* (a pesar de que en su artículo de 1974 no aparezca la palabra «*qualia*»). Incidamos de cara a explicitar de un modo conciso este vínculo entre *qualia* y conciencia fenoménica en que los *qualia* son las propiedades de los estados mentales conscientes en virtud de las cuales hay algo que es como estar en

[7] Así, por ejemplo, Marina Rakova señala en su diccionario de filosofía de la mente que los estados mentales fenoménicamente conscientes poseen rasgos cualitativos (*qualia*) y hay por tanto "something that an experience feels like" (Rakova, 2006: 140-141).

tales estados, y en que "un estado mental es fenoménicamente consciente cuando hay algo que es como estar en ese estado" (Crane, 1995/2003: 339 de la traducción).

> The fact that an organism has conscious experience *at all* means, basically, that there is something it is like to be that organism (Nagel, 1974: 436).[8]

Pero, ¿podemos definir una a una, independiente y explícitamente, las anteriores nociones? Parece que si para definir los *qualia* echamos mano del aspecto fenoménico de la conciencia y viceversa, asociando además en ambos casos el *what it is like* a dichas nociones,[9] no podremos salir de un enrevesado carrusel de definiciones que remitirían mutua y circularmente unas a otras evocando intuitivamente el sentido de las referidas nociones antes que explicitando clara y aproblemáticamente el significado de las mismas.

> Mientras no contemos con una definición afirmativa del término "fenoménico", no nos será posible evaluar afirmaciones sobre los aspectos fenoménicos y, mientras las definiciones de "fenoménico" y "qualia" remitan una a la otra, seguiremos sin saber con exactitud de qué estamos hablando cuando nos referimos a los qualia (Dennett, 2005: 79 del original, 98 de la traducción).

El problema no parece fácil de resolver mediante definiciones explícitas, cosa que, en contraste con Dennett, no preocupa excesivamente a otros.[10] En nuestra opinión, el hecho de que la solución a dicho problema parezca evitarnos se debe a que cabe la posibilidad de que Wittgenstein (1921, §5.6) no estuviera en lo cierto y el límite de nuestro mundo se halle un poco más allá de los límites de nuestro lenguaje, y en particular porque cabe la posibilidad de que los límites de nuestra experiencia

[8] Cursivas en el original.

[9] E incluso la de "experiencia", la integración de la cual en este círculo asertivo no resulta inusual y puede conducir tanto a una mayor complejidad conceptual como una más penetrante capacidad para la evocación. "Phenomenal consciousness is experience. Phenomenal conscious properties are the experiential properties of sensations, feelings and perceptions, for example, what it is like to experience pain, what it is like to see, to hear and to smell" (Block, 1994c: 27). Block (1995; 2002) propone una definición similar. Por su parte, Carruthers (2000) ha trazado una distinción entre la noción de *qualia* y la de *what it's like*. Según su propuesta, un *quale* se presenta como parte del mundo, no como parte de la propia mente (el rojo que experimento es del tomate y el dolor que experimento es de mi pierna). Por su parte, el *what it's like* sería el modo, específicamente mental y experienciado como tal, en que se nos presentan nuestros *qualia*. Así, un *quale* podría describirse usando términos del lenguaje cotidiano mientras lo mismo no sucedería con el *what it's like*. Podemos prescindir de esta distinción dada su escasa aceptación y dada asimismo la innecesaria proliferación terminológica a que conduce.

[10] "I cannot define phenomenal consciousness in any remotely non-circular way. I don't consider this an embarrassment. The history of reductive definitions in philosophy should lead one not to expect a reductive definition of anything. But the best one can do for phenomenal consciousness is in some respects worse than for many other things because really all one can do is point to the phenomenon" (Block, 2002: 206).

consciente sean más difusos que los de nuestras estructuras conceptuales,[11] las cuales pueden comprometerse sin apenas arriesgarse a dejar parte de la realidad que sea el caso de lado con determinadas tareas, con tareas como, por ejemplo, formular y dar respuesta a preguntas tales como "¿ha excedido el mercurio del termómetro la marca número veinte?" o "¿sigue en pie la Torre inclinada de Pizza?", mientras que la tarea de ofrecer respuesta a preguntas tales como "¿cómo es tu experiencia visual de mi camisa verde?" o "¿de qué modo experimentas la fiebre?" parece apuntar a un lugar en el que la frontera entre lo expresable y lo –quizá– inefable aparece desdibujada. No ahondaremos en este punto, sino que nos conformaremos con señalar a este respecto que tal vez el escasamente definido trazado de estas demarcaciones explique el hecho de que todas las definiciones de «*qualia*» que pueden encontrase en bibliografía disponible se hallen antes dirigidas a nuestra intuición que a nuestra capacidad para operar con conceptos.

De vuelta a nuestro esquemático recorrido por la breve historia del intento de definir de la noción de *qualia*, nos detendremos brevemente en otro punto de referencia fundamental dentro del actual debate: el artículo de principios de los ochenta "Epiphenomenal qualia", de Frank Jackson, otro de los hitos que jalonan la señalada historia. En este artículo Jackson trata de utilizar la noción para refutar el fisicalismo mediante el famoso experimento mental de Mary, la neurofisióloga del color forzada a una existencia acromática. Discutiremos el contenido del mismo al ocuparnos de los argumentos misterianos al final de esta primera parte. Aquí sólo nos interesa pasar sumariamente revista del modo en que los *qualia* han sido conceptualizados a lo largo de su andadura en el debate contemporáneo en torno al problema de la conciencia. A este fin, basta indicar que Jackson trata en este influyente artículo a los *qualia* como ciertos rasgos cualitativos de las sensaciones corporales y de determinadas experiencias perceptuales que resultarían, en todo caso, inaprensibles a través del mero acopio de información física.

La discusión contemporánea sobre los *qualia* gira en gran medida, como hemos destacado ya, en torno a la forma correcta de caracterizarlos. Un extremo crucial en este sentido es el de la extensión del término: ¿qué estados mentales entenderemos como portadores de *qualia*? Acabamos de mencionar la caracterización que ofrece Jackson de los *qualia* como rasgos cualitativos de sensaciones corporales y experiencias

[11] Strawson defiende una idea similar: "experiential phenomena outrun the resources of human language" (Strawson, 1994/2010: 62).

perceptuales. Puede apreciarse que el rango de estados mentales a los que así estaría atribuyendo *qualia* es ciertamente limitado.[12] Otros autores, como Block, Flanagan, Searle o Strawson, defienden caracterizaciones más amplias, y algunos llegan a plantear que todos los estados mentales conscientes poseen *qualia*. Veamos algunos ejemplos (tomados, respectivamente, de textos de Block y Searle):

> Qualia are experiential properties of sensations, feelings, perceptions and, in my view, thoughts and desires as well (Block, 1994b: 514).

> «Conciencia» y «qualia» son simplemente conceptos de igual extensión (Searle, 2007c: 98 del original, 123 de la traducción).

Si pretendiéramos extraer de este breve recorrido a través de los diferentes conatos definitorios de los que ha sido objeto la noción de *qualia* una caracterización general de la misma, el camino más directo pasaría por la acostumbrada referencia a propiedades cualitativas o fenoménicas de determinados estados mentales en virtud de las cuales cabe decir que hay algo que es como tener o estar en esos estados mentales, es decir, propiedades que consisten en cómo es tener o estar en esos estados mentales. Estas propiedades, como apuntábamos, suelen caracterizarse negativamente, contraponiéndolas a los aspectos intencionales o representacionales de los estados mentales, con lo cual se pretende dejar sólo espacio para que los *qualia* salgan a escena engalanados con los atavíos propios de eso que los filósofos de la mente denominan aspecto cualitativo de las experiencias conscientes, ése aspecto que determina *cómo* o *como-qué* es para el individuo tener tales o cuales experiencias. Pero, ¿cuáles son los estados mentales dotados de dicho aspecto? Como indicábamos, éste es un extremo ciertamente conflictivo. Casi todos los autores involucrados en el debate admiten que las experiencias perceptivas, las sensaciones corporales, las reacciones afectivas o emocionales y los estados anímicos son clases de estados mentales que, efectivamente, poseerían *qualia*. Muchos opinan asimismo que las sensaciones corporales (táctiles, térmicas, nociceptivas, etc.) poseen de hecho sólo aspectos cualitativos (*qualia*) mientras carecen de contenido intencional. Son, en cambio, algunos menos los que incluyen en el catálogo de los estados mentales poseedores de *qualia* a las experiencias

[12] Es frecuente encontrar que los *qualia* se definen de este modo restringido en someras caracterizaciones generales (vid., v. g., la entrada "*qualia*" en el glosario incluido en Blackmore, 2006: 266 del original, 362 de la traducción).

de pensamiento, como entender o recordar súbitamente,[13] así como lo son los que afirman que los estados de actitud proposicional portarían determinados *qualia*.[14] Por último, Searle (2000a; 2004a; 2007c) ha defendido, como veíamos, que las nociones de conciencia y *qualia* son coextensivas, dado que no hay experiencia consciente sin aspecto cualitativo, llegando a plantear así que el concepto de *qualia* es engañoso, ya que parece sugerir que algunas experiencias conscientes carecen de aspecto cualitativo. Volveremos sobre este punto al final de este capítulo.

Es importante, por otra parte, no perder de vista que no hablamos de entidades cuando nos referimos a los *qualia*, ya que no existe ninguna entidad independiente que responda o pueda responder a ninguna de las definiciones de *qualia* que hemos expuesto hasta ahora, y esto porque los *qualia* no son sino un aspecto de determinadas entidades peculiares y difíciles de caracterizar en sí mismas: los estados mentales conscientes. Del mismo modo que parece innegable que la masa y la forma del planeta Tierra existen sin ser por ello entidades autónomas, discretas o independientes, asimismo los *qualia* parecen existir como propiedades de determinados estados mentales. Así, cuando hablamos de *qualia* no estamos refiriéndonos a entidades, sino a propiedades de los estados mentales conscientes. Por tanto, no cabe hablar, *stricto sensu*, de la existencia o inexistencia de los *qualia*, sino sólo de la pertinencia de atribuir a determinados estados mentales determinadas propiedades.

Sentado pues que hablamos de propiedades, cabe señalar que las mismas son habitualmente concebidas como poseedoras, a su vez, de una serie de propiedades –propiedades de propiedades, pues, esto es, propiedades de segundo orden– acerca de las cuales reinan el debate y las discrepancias. También en este punto nos encontramos con que no existe acuerdo acerca de la forma adecuada de caracterizar a los *qualia*. Cuáles de entre las propiedades de segundo orden que podemos encontrar en la bibliografía sobre ellos es pertinente atribuirles es un tema actualmente en litigio y, ciertamente, la discusión acerca de este extremo no tiene visos de ir a resolverse mediante un *experimento crucial* o una inconcusa elucidación acerca de la cual no pueda caberle a ningún especialista hesitación de ninguna clase: es verdaderamente difícil figurarse el modo en que el consenso pueda alcanzarse en este punto. Así las cosas, no cabe ninguna

[13] Strawson (1994/2010) puede ser considerado el abanderado de quienes así lo hacen.

[14] Flanagan (1992, cap. 4) ha planteado que los estados de actitud proposicional son cualitativos, planteamiento al que teóricos como Lormand (1995; 1998: 117) han venido oponiéndose dado que consideran que son los *qualia* de los estados perceptivos o anímicos que acompañarían a los estados de actitud proposicional los que aportan ese componente cualitativo, el cual, desde esta perspectiva, desaparecería si pudiera aislarse una actitud proposicional disociándola de todo estado mental cualitativo concomitante.

manera de presentar la lista de propiedades de segundo orden que entendemos pertinente atribuir a los *qualia* que exceda la exposición de los motivos por los cuales rechazamos alguna de las propiedades a menudo atribuidas a los *qualia* en la bibliografía, así como la de aquéllos por los cuales nos adherimos a la serie de propiedades de segundo orden que, en efecto, entendemos que cabe atribuirles.

Nos centraremos a continuación en las principales de entre estas propiedades de segundo orden, criticando con algún detalle la primera (la supuesta propiedad de la intrinsecalidad) y exponiendo el sentido en que entendemos que los *qualia* poseen el resto. Rechazaremos pues la idea de que los *qualia* son intrínsecos y los caracterizaremos como privados, inefables, directamente accesibles y exclusivos de los estados mentales conscientes.

a) Los *qualia* son intrínsecos. Esta supuesta propiedad de los *qualia* pretende presentárnoslos como aspectos no relacionales de los estados mentales. Adoptando la terminología funcionalista al uso, podría decirse que, según la definición habitual de esta propiedad de los *qualia*, éstos son independientes de las cadenas causales o funcionales entre inputs, estados mentales y outputs. En otras palabras, según esta propiedad de la intrinsecalidad, los *qualia* serían propiedades que los estados mentales tienen en sí mismos y no en virtud de su contenido u objeto intencional, su causa externa o interna o su relación con otros estados mentales, mas, la habitual estrategia consistente en tratar de demostrar mediante experimentos mentales tal independencia del aspecto cualitativo o fenoménico de los estados mentales no ha resultado por el momento concluyente. En este sentido, será de utilidad el siguiente ejemplo a modo de epítome: la física muestra un nada desdeñable catálogo de propiedades existentes pero no autónomas o causalmente aisladas, como la presión, la temperatura y el volumen, la velocidad, la aceleración y la inercia, la velocidad y la temperatura, la dureza, la tenacidad, la plasticidad o la resiliencia, mas nadie pretendería demostrar la existencia de alguna de estas propiedades recurriendo, digamos, a estrategias *cæterīs pāribus*, porque de hecho nadie ha dado por lo pronto con una propiedad independiente de todo el resto de las propiedades, condiciones o acontecimientos. ¿Qué nos invita, pues, a pensar en la existencia de propiedades que pueden permanecer constantes mientras varía absolutamente todo a su alrededor o, viceversa, variar mientras todo permanece constante? Por ejemplo, si hemos de entender que intrínseco significa independiente de la relación entre el sistema visual y el objeto causante de una percepción (incluso de

meras relaciones dentro del propio sistema nervioso, si es que pretende el defensor de la intrinsecalidad contraatacar con el argumento de la ilusión), evidentemente no aceptaremos tal propiedad como propiedad de los *qualia*. El modo en que experimento el naranja de esa naranja podría concebirse como lógicamente independiente de las propiedades reflectantes de su superficie, así como de la actividad de los conos de mis retinas y el resto de procesos neuronales, desde el nervio óptico, el quiasma óptico, el tracto óptico y los núcleos geniculados laterales hasta mi corteza visual secundaria, en las regiones superiores de mi lóbulo occipital y las posteriores de mi lóbulo parietal. No obstante, un mundo sin naranjas parece de igual modo lógicamente concebible, pues no hay –o no parece haber– en el acto de concebirlo contradicción lógica de ningún tipo, pero no es éste el mundo del que hablamos: nos conciernen las cosas tal y como son en este mundo y no tal y como cabría pensar que pudieran ser en este o aquel mundo posible. En definitiva, no tenemos ningún motivo para interpretar esta propiedad de este modo: nada sugiere que el modo en que veo el naranja de esa naranja sea independiente del modo en que la luz reflejada por su superficie afecta a mi sistema nervioso y todo apunta, de hecho, en la dirección contraria. Con todo, el modo en que veo el naranja de esa naranja puede ser entendido como una propiedad intrínseca de mi estado mental en tanto tal modo, a pesar de depender de los señalados fenómenos acaecidos entre la superficie de la naranja y mi sistema visual, forma parte de mi estado perceptivo actual, y dentro de él, lo que podríamos denominar propiedades fenoménicas relacionales[15] ("el naranja de esa naranja me resulta más parecido a este rojo que a aquel rosa") que tal modo, tal vez idiosincrásico, de ver el naranja de la naranja posee, no tienen necesariamente que guardar relación directa con sus concomitantes propiedades semánticas, es decir, con el contenido proposicional que pudiera acompañar a mi estado perceptivo. Uno podría pensar al percibir la naranja en cuestión: "una naranja de un bonito color naranja", a pesar de que las condiciones de iluminación sólo le permitieran verla en un tono muy distinto (podría uno pensar tal cosa, por ejemplo, si hubiera visto la naranja en otras condiciones de iluminación y recordara el color experimentado en aquella ocasión), o incluso podría percibir exactamente el mismo tono de naranja sin reconocer en absoluto la fruta en cuestión. «Intrínseco» vendría a significar bajo esta interpretación "relativamente independiente de la capacidad de operar con conceptos". Pero el sentido en el cual se presenta como problemática esta propiedad de la

[15] Lo que Shoemaker (1994b) ha denominado *beliefs about differences in how things appear*.

intrinsecalidad es, con todo, de grano más fino: no puede decirse que propiedades cualitativas como las señaladas sean intrínsecas en el sentido de su total independencia respecto del resto de la economía mental, ya que, de hecho, datos empíricos avalan la falsedad de esta tesis. Así, por ejemplo, los miembros de tribus no familiarizadas con la representación bidimensional de estructuras tridimensionales no perciben dichas representaciones del mismo modo que los sujetos acostumbrados a semejante tipo de representación (vid., v. g., Chalmers, 1976/2013: 6 del original, 42 de la traducción), es decir, carecen de los hábitos mentales[16] (inconscientes y adquiridos) que pudieran conformar el contexto necesario para propiciar la percepción de líneas sobre un plano como la representación gráfica de un objeto tridimensional. En conclusión, cabe dudar que los *qualia* posean la propiedad de la intrinsecalidad en el sentido de ser ellos propiedades del estado mental en sí mismo y con independencia de cualesquiera relaciones con objetos externos, estados del sistema nervioso u otros estados mentales. El único modo en que entendemos que dicha propiedad de segundo orden pudiera resultar aceptable es el ya apuntado: la relativa independencia de los *qualia* respecto de contenidos representacionales de alto nivel (semánticos): dos personas pueden mirar el mismo objeto reconociéndolo sólo una de ambas; así, es de suponer que ambas experimentan formas y colores similares, pero una de ellas no asocia esas formas y colores con ninguna entrada en su lexicon, es decir, no puede identificar el objeto. Pero parece que este tímido sentido atenuado de la noción de intrinsecalidad como relativa autonomía respecto de contenidos proposicionales no es el que propugnan los defensores de dicha propiedad de segundo orden. No obstante, incluso en este sentido atenuado, la propiedad de la intrinsecalidad sigue resultando problemática: esta independencia del aspecto fenoménico respecto del conceptual, ¿puede concebirse como absoluta? ¿No cambiará la propia experiencia perceptiva de la persona que no identificaba el objeto cuando, súbitamente, lo reconozca? Si no estuviéramos dispuestos a admitir que lo que ha cambiado –al menos en parte– con el súbito reconocimiento del objeto sea la propia experiencia perceptiva, ¿de dónde colegimos semejante independencia entre la fenomenología y la semántica de uno y el mismo estado mental? ¿Cómo demostrar que con la súbita identificación del objeto sólo cambia determinado

[16] Tal vez resulte menos comprometido expresar esta idea –haciendo caso omiso de la *falacia mereológica* (Bennett & Hacker, 2003)– como sigue: sus sistemas nerviosos no han sido convenientemente moldeados por la experiencia para percibir volumen en la apropiada disposición de líneas trazadas sobre superficies planas.

aspecto representacional de alto nivel sin variación alguna en lo que al aspecto fenoménico del estado mental, concebido como un todo, toca?

Además, esta propiedad se define habitualmente con una concisión que parece obedecer a las dificultades que entraña explicitar el sentido de la escueta glosa "los *qualia* son propiedades intrínsecas de los estados mentales conscientes, es decir, propiedades *no relacionales*". A primera vista se trata de un enunciado significativo, pero, ¿qué quiere decir "no relacional"? ¿Pueden las ciencias naturales estudiar propiedades "no relacionales"? ¿Cabe concebir algún objeto, acontecimiento o propiedad "no relacional"? Bien es cierto que los filósofos de la mente parecen caracterizarse por una insólita capacidad para concebir las escenas más inverosímiles, pero sustituyamos en el anterior interrogante «concebir» por «probar la existencia de». ¿Qué tenemos? Todas las papeletas para un rotundo no.

"No relacional" parece deslizarse subrepticiamente hacia márgenes que traen a las mientes "epifenoménico", pues algo no relacional no puede sino ser algo inerte causalmente y aislado por tanto del orden causal natural. Pero, ¿en qué sentido cabe hablar de epifenómenos o hechos o propiedades carentes de relaciones? Mi sombra existe independientemente de mi capacidad de montar en bicicleta. Es un epifenómeno desde este punto de vista: no interviene en la ocurrencia ni ocupa lugar en la explicación de mi actualización o puesta en práctica de dicha capacidad. No obstante, mi sombra no es un epifenómeno, y esto por el simple hecho de que, al igual que el resto de lo existente, es a la vez causada y causa. Mi sombra, de hecho, no es independiente de mi capacidad de montar en bicicleta: cierto que su existencia no depende de esta habilidad, pero cuando la ejercito, mi sombra presenta una serie de poderes causales que varían en virtud, precisamente, de tal actualización de dicha habilidad. También el ruido del famoso ejemplo de Huxley es epifenoménico en este sentido –es decir, es un epifenómeno desde un cierto punto de vista–, pero en ningún caso un epifenómeno en el oscuro sentido de carecer de potencia causal –pues, en este sentido, ningún epifenómeno parece tener cabida en el mundo natural, y si algo similar existiera, si algo cuya característica fundamental fuera su absoluta inercia causal hubiera por algún caprichoso albur caído en el reino de lo existente, su existencia estaría más allá de nuestra incumbencia: algo sin ningún poder causal restaría, por definición, aislado y ni tendríamos noticia de ello ni nos sería por tanto dable pensar o hablar sobre ello.

> No veo la necesidad de postular la existencia de propiedades intrínsecas en absoluto (…). Todo lo que podemos conocer son propiedades relacionales. Decir que los *qualia* son la excepción a esta regla (…) es sólo una expresión de deseo que no prueba que los *qualia* sean propiedades intrínsecas (Pérez, 2002: 77-78).

Decir que los *qualia* podrían de hecho existir qua propiedades de cierta clase de estados mentales y de hecho hallarse inextricablemente relacionados con los aspectos intencionales o representacionales que los mismos pudieran poseer y con las estructuras, redes o cadenas causales o funcionales que describirían –en un contexto científico ideal– la actividad de nuestra mente-cerebro, ¿tiene necesariamente que significar que carecen de cualquier rasgo que exceda lo representacional, intencional o funcional? La mayor parte de los experimentos mentales relacionados con el problema de los *qualia* parecen estar diseñados para demostrar que éstos son de hecho arrelacionales o independientes de cualesquiera aspectos intencionales o representacionales o cualesquiera estructuras, redes o cadenas causales o funcionales. Pero, ¿han de implicarse esos extremos? ¿No podría darse el caso de que los *qualia* fueran propiedades de nuestros estados mentales pero que no existieran con total independencia de cualesquiera aspectos intencionales o representacionales o cualesquiera estructuras, redes o cadenas causales o funcionales ni hubieran de describirse como exhaustivamente constituidos por rasgos intencionales o representacionales? Como veremos en nuestra crítica de los argumentos misterianos –y desarrollaremos en el segundo capítulo de la segunda parte–, nadie ha logrado obturar esta vía.

Apuntemos para terminar que al plantear el tipo dudas que suscita la supuesta intrinsecalidad de los *qualia* es frecuente topar con el siguiente tipo de réplica: algunas propiedades, como "ser mayor que", son constitutivamente relacionales: su definición depende del establecimiento de relaciones. Otras, como los *qualia*, o "ser un cuadrado", no, dado que podemos especificar qué son y determinar así su estatus ontológico sin hacer mención alguna de sus relaciones con otros hechos o propiedades. Así, se dice, en el caso de un cuadrado real, siempre necesitaremos para definirlo de forma acabada hacer mención de su color, sus relaciones con otros objetos y propiedades, etc. Sin embargo, especificar qué es "ser un cuadrado" no requiere de mención alguna de relaciones con otras propiedades u objetos. Bien, sólo cabe responder que si la idea es platonizar los *qualia*, entonces la idea platónica de espermátida es tan intrínseca como la de *qualia*. En cambio, los *qualia* con los que aquí nos las habemos y los *qualia* a los que los "qualófilos intrinsecalistas" se refieren como intrínsecos han de ser, al igual que los

cuadrados reales, y en tanto fenómenos naturales, tan arrelacionales como las espermátidas reales (si es que el matiz tiene algún sentido, lo que dependería de la existencia de una idea de espermátida o espermátida ideal a la que contraponer las espermátidas reales). Si los *qualia* son intrínsecos en el sentido que esta réplica pretende, es decir, en el sentido de la posibilidad de especificar su estatus ontológico sin mención alguna de su relación con otros hechos o propiedades, entonces, potencialmente, todo podría serlo, dado que, a pesar de su efectiva dependencia respecto de otros fenómenos o propiedades, siempre podríamos improvisar una definición que presente cualquier propiedad u objeto como autónomo y arrelacional. Radicalizando nuestra contrarréplica, cabe decir que precisamente eso es lo que hace Platón con la idea de identidad en el Timeo: sustantiviza un adjetivo e improvisa la propiedad intrínseca de "ser idéntico a" bajo la denominación «lo idéntico». Ahí tenemos una propiedad relacional intrínseca: hasta ellas pueden platonizarse si es eso lo que se desea.

Pero cabe también la posibilidad de que hayamos descuidado algo fundamental y especialísimo que sólo los *qualia* poseen. No la descartamos. Sin embargo, no encontramos una interpretación más afortunada de este tipo de réplica: los *qualia*, según la misma, no necesitan para ser especificados ninguna de sus relaciones –posibles, empíricas, efectivas, factuales, no lógicas– y son así intrínsecos. La empírea espermátida ideal tampoco necesita para ser especificada ninguna de sus relaciones contingentes, mundanas, efectivas, concretas, positivas, pero la espermátida ideal no preocupa demasiado a los embriólogos, acostumbrados a considerar sus objetos de estudio dentro de un marco metodológico, epistemológico y ontológico un tanto distante del platónico, el de los fenómenos naturales, entre los cuales no contamos ni parece que vayamos a contar nunca con un corpus demasiado definido y abundante (Ø, exhaustivamente) de propiedades intrínsecas.[17]

b) Los *qualia* son privados. Esta propiedad de los *qualia* no parece objetable más que desde un punto de vista radicalmente operacionalista o verificacionista. Ciertamente, ninguno de los hechos estudiados por las ciencias naturales posee esta característica de la privacidad. Además, la replicabilidad y publicidad son consideradas a menudo elementos indispensables del método científico, circunstancias que harían de

[17] Destaquemos para terminar que la bibliografía filosófica disponible acerca de la caracterización de la propiedad de la intrinsecalidad –vid., v. g., Lewis (1983b; 1986), Sider (1996), Vallentyne (1997), Yablo (1999)– no alcanza a resolver las dudas suscitadas por el uso que los filósofos de la mente hacen de dicha noción.

los *qualia* un nada habitual objeto del estudio científico. No obstante, esto no es ningún argumento contra la efectividad de la existencia del señalado modo, tal vez idiosincrásico, en que yo percibo un determinado tono naranja, ni contra mi incapacidad para percibirlo del modo en que otro u otros pudieran percibirlo. Los *qualia* serían pues propiedades de los estados mentales conscientes dadas de forma particular y exclusiva al sujeto, es decir, al organismo que los instancie.

c) Los *qualia* son inefables. Esta propiedad de segundo orden se halla estrechamente relacionada con la anterior.[18] Siendo los *qualia* propiedades de la experiencia consciente que permanecen en el ámbito privado, es decir, que no son susceptibles de ser compartidas ni exhibidas públicamente, parece que nos hallamos con ellos ante el escarabajo en la caja de Wittgenstein (1953: §293). ¿Qué tipo de lenguaje podemos emplear para describir algo que tan siquiera podemos asegurar que compartamos? Los *qualia* son propiedades de los estados mentales conscientes inexpresables o, cuando menos, no exhaustivamente comunicables. Al igual que la anterior, esta propiedad de los *qualia* vuelve a mostrárnoslos como perspectivos, lo cual quiere en este caso decir que con independencia de la cantidad y asimismo la calidad de las palabras que mi interlocutor utilice, si no he padecido nunca dolores renales –si no puedo adoptar sobre ellos la perspectiva del paciente, si la misma me resulta totalmente ajena– al menos una parte del referente de su discurso permanecerá para mí inaccesible cuando me hable de ellos (máxime si yo nunca hubiera experimentado dolor de ninguna clase), dado que en el caso de conceptos fenoménicos (como "dolor" o "dolor renal")[19] el significado está parcialmente constituido por la propia experiencia: es necesario haber experimentado determinado tipo de *qualia* para comprender qué es como tenerlos y, por tanto, para comprender cabalmente a qué se refieren los conceptos fenoménicos asociados a los mismos. El sentido en que a los conceptos fenoménicos les cabe fungir como atenuantes de la inefabilidad de los *qualia* puede oscurecer la idea que se esconde tras esta propiedad. Considérese la comunicación sobre sabores entre dos personas sin ningún problema perceptivo. Independientemente de la cantidad y la calidad de las palabras utilizadas, el modo en que uno experimenta un determinado sabor, ¿puede, aun cuando los interlocutores compartan los conceptos fenoménicos pertinentes, ser

[18] Con «inefable» no pretendemos significar aquí "netamente inexpresable", sino más bien "no exhaustivamente caracterizable haciendo uso de cualesquiera recursos de nuestro lenguaje".
[19] Para una introducción a esta noción –usada por primera vez en Loar (1990)– véase Balog (2009).

enteramente comunicado? La inefabilidad que hemos presentado como propiedad de los *qualia* parece pues, en el fondo, más bien una propiedad de nuestro lenguaje.

El crítico de esta propiedad de segundo orden de los *qualia* se vería en la necesidad de demostrar que el modo en que ve, huele o saborea las cosas (e incluso el modo en que su hijo desea una bicicleta, en el caso de que estemos dispuestos a atribuir carácter fenoménico a estados de actitud proposicional) puede ser comunicado de forma exacta y precisa, de tal manera que cualquier hablante competente de su idioma –v. g., un ciego de nacimiento interesado en conocer la diferencia entre "magenta" y "púrpura"– ha de poder comprender exhaustivamente sus descripciones.

d) Los *qualia* son directamente accesibles. Al tratarse de propiedades de los estados mentales vinculadas al particular modo en que experimentamos los diferentes aspectos de nuestra vida mental, es decir, propiedades de nuestros estados mentales que no necesitan para su actualización de derivación mediante razonamiento o inferencia, ningún tipo de operación mental es necesaria para el acceso a ellos –y bien cabe que esta extendida metáfora del acceso resulte espuria: la experiencia se nos da, es lo dado, y no es sencillo determinar a través de qué derrotas se supone que "accedemos" a ella–. Son los *qualia*, en este sentido, propiedades primarias o directas de la experiencia, es decir, propiedades que no se coligen de otros contenidos.

e) Los *qualia* son propiedades exclusivas de los estados mentales conscientes. Ningún estado mental no consciente posee tales propiedades. Además, para todo estado mental consciente –considerado como un todo– hay algo que es como estar en él. Este es un punto crucial, dado que admitido esto no podremos asentir ante la frecuente afirmación según la cual las actitudes proposicionales sólo tienen contenido intencional. Puede que "en sí mismas" se limiten a representar el mundo de determinada manera, pero, por ejemplo, una creencia nunca flota en el vacío, y no podemos creer que llueve sin creerlo de determinada manera (sorprendidos, molestos, alegres). Además, sin asociar la creencia a otros aspectos concomitantes de los estados mentales y considerándola "en sí misma", la misma creencia puede experimentarse de distintos modos: puedo creer que lloverá estando totalmente convencido de ello (puedo tener incluso la certeza de que así será) y manteniendo dicha creencia en el mismo centro de lo que James denominara foco atencional, o puedo creerlo sin reparar demasiado en ello ni darle mucha importancia mientras el núcleo de dicho foco lo ocupa, por ejemplo, mi

mano izquierda, incapaz de cambiar a tiempo de acorde al llegar a determinado punto de la partitura. Cabría acentuar que, adicionalmente, cada clase de estado de actitud proposicional se experimenta de un modo diferente: imaginemos, si nos es posible, un estado mental aislado –sin percepciones ni estados anímicos concomitantes–, e imaginemos que se trata de un estado de actitud proposicional cuyo contenido es "hay millones de euros en mi cuenta bancaria". ¿No hay, acaso, algo que es como creerlo y algo, diferente, que es como desearlo?

Nos inclinamos, pues, con Block, Flanagan, Searle o Strawson, por una concepción amplia de la extensión de «*qualia*» según la cual ésta coincide con la de «experiencia»,[20] que a su vez coincide con la de «conciencia fenoménica». Tanto recordar como dudar, creer, valorar, oler, doler o añorar son modalidades de experiencia que pueden preponderar unas sobre otras, pero nunca darse totalmente aisladas. "No hay una sensación, y luego una percepción, y luego una cognición y luego una emoción" (Vilarroya, 2002a: 263). Puede que, en un momento dado, una homogénea experiencia visual del cielo despejado ocupe el núcleo del referido foco atencional jamesiano de algún sujeto, y puede que el mismo roce un singular y uniforme trance vipassánico, pero, si no es una cámara de vídeo, esa experiencia estará sin duda acompañada por una dinámica cohorte fluctuante en la periferia del asimismo fluctuante foco atencional, una comitiva integrada por sensaciones corporales, reminiscencias en forma de imágenes acústicas, un determinado tono emocional, fragmentos de *inner speech*… una comitiva cuyos integrantes alzan constantemente unos sobre otros sus voces, hora quedas hora rotundas.

Una curiosa implicación de esta concepción de la extensión de «*qualia*» es la de la superfluidad del término. El desmesurado caudal de publicaciones sobre los *qualia* que las últimas décadas se han visto obligadas a drenar puede hacer pensar que, en efecto, el término es de algún modo indispensable para abordar el problema de la conciencia. Sin embargo, la solución del problema de la conciencia, esto es, la explicación científica de dicho fenómeno biológico –dedicamos la segunda parte, en buena medida, a hacer explícita la pertinencia de esta equiparación–, podrá habérselas perfectamente sin él, y ello a causa de su redundancia, su no dejar de constituir una forma alternativa de aludir al *explanandum* de la apuntada explicación.

[20] "It's not like there are experiences of kind A that are or have qualia and experiences of kind B that do not" (Hales, 2010: 24-25).

*Hasta aquí, nos hemos encargado de presentar el contexto en que se enmarca el estudio contemporáneo de la conciencia. La introducción de la contraposición entre lo fenoménico y lo intencional resulta indispensable para la adecuada comprensión de ese contexto. En este capítulo la hemos introducido de forma expositiva, definiendo sus términos y denunciado la inconsistencia de la noción consuetudinaria de uno de ambos. Hemos tratado, en cualquier caso, de cada uno de esos términos por separado. Más adelante, en la segunda parte, una vez presentado el modo en que las diferentes teorías de la conciencia de las que nos ocuparemos en los dos próximos capítulos hacen uso de esos términos y trazan vínculos entre ellos, nos encontraremos en situación de argumentar sobre las concepciones en liza acerca de la relación entre los mismos y sobre lo escasamente provechosa que desde el punto de vista heurístico, metodológico y, en general, epistemológico viene resultando la confrontación entre dichas concepciones. Por lo pronto, detengámonos en esas teorías. Dedicaremos el próximo capítulo a las ontológicas y el siguiente a las explicativas.*

# CAPÍTULO 4

## DÓNDE ENCAJA LA CONCIENCIA. MAPA DE LAS PROPUESTAS ONTOLÓGICAS

### 1._Esquema de las posturas tradicionales: una breve prehistoria del problema de la conciencia

Como una gran cantidad de autores ha sugerido desde diversos ángulos y disciplinas (vid., v. g., Alter & Howell, 2011; Bruce Goldstein, 2010: 39; Carrier & Mittelstrass, 1995: 180; Chalmers, 1996: apdos. 1.4 & 5.3; Fernández-Guardiola, 1979; González Álvarez, 2010: 267; Harris, 2006: 1-4; Hutto, 2000: 5; McGinn, 1991: 1; Nagel, 1993b; Pawlik, 1998; Searle, 1992: 100 del original, 112 de la traducción; 1993a: 8; 1998: 379; 2002b: 57; Solms & Turnbull, 2002: cap. 2; Thagard, 2005: 175 del original, 265 de la traducción; Van Gulick, 2001: 1; Velmans, 2009: 3; Vicari, 2008: 1; Zangwill, 1977[1]), el problema de la conciencia, planteado en términos contemporáneos, puede ser entendido como una reformulación del tradicional problema mente-cuerpo (habitualmente remozado como "problema mente-cerebro"). Igualmente, no han sido pocos los que han incidido en la antigüedad del problema. Algunos (v. g., Martínez-Freire, 1995: 123-125; 1999: 67; 2007b: 799; Van Gulick, 2011) han llegado a sugerir que se trata de un problema acerca del cual ha reflexionado la humanidad desde sus albores hasta nuestros

[1] Richard L. Gregory ha escrito recientemente (Gregory, 2001) una breve e interesante reseña biográfica del poco recordado neuropsicólogo británico autor del texto de esta última referencia, un texto que avanzaba en una época temprana un polifronte interés por la conciencia desde la psicología, las neurociencias y la filosofía, un texto, al igual que la referida semblanza, breve y asimismo plagado de referencias interesantes: en pocas páginas recorre con solvencia un trecho que va de Fechner a los cerebros divididos que en aquella época estudiaba junto a un joven Michael S. Gazzanigga Roger W. Sperry –quien, por cierto, citara a Zangwill en un artículo (Sperry, 1984) sobre conciencia, identidad y cerebros divididos publicado poco antes de la muerte de éste.

días. Otros, partiendo de investigaciones antropológicas, se han limitado a señalar que las prácticas funerarias neolíticas evidencian –aunque indirectamente, pues lo que evidenciarían directamente sería, cuando más, la presencia de cierta suerte de creencias de carácter espiritual o religioso– al menos cierto grado de pensamiento reflexivo acerca de la naturaleza de la conciencia (vid. Pearson, 1999; Clark & Riel-Salvatore, 2001). En cualquier caso, la antigüedad de la reflexión acerca del problema mente-cuerpo puede columbrarse en el hecho de que su tratamiento resultara de algún modo ineludible para quienes redactaran los textos religiosos más antiguos de los que tenemos noticia.[2]

No obstante, no sería hasta la germinación del fermento griego de nuestra cultura occidental que se dieran los primeros pasos hacia la formulación del problema en términos similares a ésos en que hoy es discutido. Así, ya en las obras de Homero y Hesíodo pueden intuirse algunos de los contornos de lo que posteriormente vendría a conformar el núcleo del problema mente-cuerpo. En ellas, la noción de mente o alma (ψυχή) aparece como contrapuesta a la de cuerpo (σῶμα), siendo éste concebido como mera materia inerte que sólo en virtud de aquélla accede al reino de la orgánica, esto es, adquiere vida. De este modo, al abandonarlo aquélla, es decir, con la separación del cuerpo y el alma o mente, sobreviene la muerte. Posteriormente, a partir del siglo VI a. C., los primeros planteamientos filosóficos sobre las relaciones entre el cuerpo y la mente ofrecen una serie variaciones sobre los planteamientos homéricos. Los primeros filósofos griegos mantuvieron la concepción homérica del alma o mente como principio vivificador, a la base de los procesos que intervienen en la génesis, el desarrollo y la realización de las funciones de los seres vivos, y atribuyeron a dicho principio una naturaleza material, aunque entendieran que la materia que constituye el alma es de una clase distinta, más tenue o sutil que esa otra que integra los cuerpos. Esta concepción, que podemos denominar *monismo sustancialista antiguo*, se presentó en la filosofía griega anterior a Sócrates en diversas variantes, como el alma constituida por aire de Anaxímenes (vid. fragmento 2, en Bernabé Pajares, 2001: 65) o el alma integrada por átomos ígneos de Demócrito (vid. Aristóteles, *De Anima*, I, 2, 404 a 1-10).[3] Hacia esa misma época, ideas

[2] Estos textos serían *Los Textos de los sarcófagos*, inscritos durante el Imperio Medio de Egipto (entre 2300 y 2000 a. C) pero provenientes de los *Textos de las pirámides*, grabados durante el Imperio Antiguo (entre 2500 y 2300 a. C). En el núcleo de las ideas en los mismos contenidas se hallan las de la inmortalidad y la resurrección en la forma en que las mismas llegaran al *Libro de los muertos*, ideas que suponen una cierta elaboración de una determinada concepción acerca de las relaciones entre el alma y el cuerpo.

[3] José Luís González Recio ha propuesto recientemente que, al atribuir al alma este tipo de naturaleza material sutil, estarían estos pensadores asegurándose de que las almas postuladas en sus teorías pudieran resultar tan ágiles y flexibles como parece serlo el pensamiento (González Recio, 2007: 169).

religiosas[4] como las que conformaran el substrato del orfismo y el pitagorismo contribuyeron a una conceptualización diversa de las relaciones entre el cuerpo y la mente que traería consigo un distanciamiento del monismo sustancialista antiguo. Las potencias del alma son concebidas aún de forma análoga, como principio vivificador, pero su origen, su forma de existir y su naturaleza son presentadas de un modo radicalmente diferente: ahora el alma tiene un origen divino, una existencia eterna y una naturaleza inmaterial. Surge con esta nueva concepción de las relaciones entre el alma y el cuerpo un enfoque ontológico dualista que ha atravesado la historia de occidente llegando a nuestros días. La primera doctrina dualista articulada filosóficamente de que tenemos constancia se debe a Platón y puede leerse en dos de sus obras de madurez: el *Fedón* y la *República*. Platón recoge en el primer diálogo referido la tradición órfico-pitagórica de la divinidad e inmortalidad del alma y la presenta como la parte del ser humano que vincula a éste con el empíreo e incorruptible mundo de las ideas; por su parte, del cuerpo nos dice que hemos de entenderlo como una suerte de cárcel corruptible en virtud de la cual el alma permanece inmersa en el cambiante mundo sensible hasta que, con la muerte, es liberada.[5] Tanto en la *República* como, posteriormente, en el *Fedro* y el *Timeo*, Platón esboza un esquema tripartito en el cual sólo una de las partes del alma se presenta como separable del cuerpo: el segmento concupiscente del alma, origen de apetitos y deseos sensibles, y el irascible, origen de tendencias y pasiones nobles como la valentía o el arrojo, permanecen ligados al cuerpo y mueren con él, mientras el segmento racional o inteligible aparece en el planteamiento platónico como una especie de sitial de la razón desde el que deben embridarse rígidamente los dos primeros y que existirá tras la muerte de la persona del mismo modo que existió antes de su nacimiento, aunque desprovisto en ambos casos del carácter personal propio de la concepción homérica, retomada más tarde por el cristianismo. Pero, ¿cómo se relaciona esta defensa de la inmortalidad del alma con la ontología de lo mental? En la metafísica platónica lo verdaderamente real no tiene ya una naturaleza material, como sucediera en la escuela milesia y en los atomistas. Lo material se presenta como efímero, cambiante y engañoso, motivo por el cual Platón prefiere reservar ese estatus de *lo verdaderamente real* para su mundo de las ideas: los objetos materiales son meras copias imperfectas de las ideas

[4] "The distinction between mind and matter, which has become a commonplace in philosophy and science and popular thought, has a religious origin, and began as the distinction of soul and body." (Russell, 1945: 134 del original, 176 de la traducción).

[5] Platón, tomando el espíritu de los pitagóricos y la letra (la expresión) de Eurípides, llega a referirse al cuerpo como un contaminado sepulcro ambulante (*Gorgias*, 492e; *Crátilo*, 400c; *Fedro*, 250c).

inmateriales. Las ideas posibilitan tanto la existencia del mundo como nuestra comprensión del mismo, y con esto llegamos a la respuesta a la pregunta planteada, pues uno de los argumentos que Platón utiliza en el *Fedón* para defender la inmortalidad del alma tiene la siguiente forma: dado que en las ideas reside la inteligibilidad del mundo, y dado que el alma debe captar las ideas para comprender, ha de haber entre ellas una afinidad que permita al alma captar las ideas, de donde concluye que esa afinidad habrá de residir en la naturaleza inmaterial que alma e ideas compartirían.

Con Aristóteles y su abandono de la doctrina platónica de las ideas inmateriales el dualismo entre el mundo sensible y el mundo inteligible es sustituido por un dualismo conceptual basado en la distinción hilemórfica entre materia y forma. Dentro del esquema trazado por esta distinción, el problema mente-cuerpo es reelaborado desde un punto de vista inédito en la tradición previa: el alma pasa a ser definida dentro de dicho esquema como la forma del cuerpo. Sus propiedades, sin embargo, son descritas de forma muy similar a la habitual en los siglos anteriores, esto es, como principio vivificador y regulador de las funciones vitales. Como Platón, Aristóteles distingue tres partes o –resulta en su caso más apropiado decir– tipos de alma: la vegetativa, que anima a los vegetales causando su génesis y desarrollo; la sensitiva, presente en los animales y que a las funciones de la anterior añade las motrices y perceptivas; y la racional, que, presente sólo en el ser humano, viene a sumar a las potencias de las anteriores la del razonamiento. Así, alineándose con la tradición, Aristóteles atribuye al alma la capacidad de sustentar y posibilitar el despliegue de los procesos vitales e intelectuales. No obstante, la concepción aristotélica de las relaciones entre el alma y el cuerpo presenta un carácter novedoso, quizá menos "trascendental" que el dualismo platónico, pero también más equívoco: a día de hoy la ontología aristotélica de lo mental sigue abierta a interpretaciones a causa del modo en que en diferentes lugares de su obra se alcanzan conclusiones difícilmente conciliables. La materia es para Aristóteles nuda potencia: algo que puede ser cualquier cosa, algo netamente indeterminado. Pero toda indeterminación ha de ser determinada, y todo lo existente está así necesariamente compuesto de una materia determinada por una forma. Materia y forma, de este modo, no pueden existir por separado y, por ende, tampoco el alma, forma del cuerpo, puede existir separada de éste. Con todo, la filosofía de la mente de Aristóteles sigue, como apuntábamos, abierta a interpretaciones, dado que, a pesar de que su hilemorfismo excluye, en principio, la posibilidad de concebir el alma como separable del cuerpo, Aristóteles no parece acabar de exorcizar con él toda forma de dualismo, pues insiste en presentar al intelecto

o parte racional del alma como una facultad excepcional de ésta, señalando que, después de todo, podría ser separable del cuerpo (*De Anima*, II, 1, 413 a 6-7) y carecer de un órgano corporal (*De Anima*, III, 4, 429 a 10 – b 9).[6]

En el tránsito de la filosofía clásica a la medieval la especulación acerca de las relaciones entre la mente y el cuerpo, entre el alma y el mundo material, cobraron un cariz enteramente dominado por el influjo de la religión cristiana, a la luz de la cual se sucedieran los conatos de relectura catequizante de las fuentes platónicas y aristotélicas. Al igual que el resto de lo existente, y al contrario de lo que sucedía en la ontología griega, el alma de todo ser humano es según la concepción cristiana producto de la creación divina, esto es, al igual que el resto de lo existente, el alma de todo hombre es creada de la nada y tiene por tanto un comienzo concreto. Además, toda alma es fruto de una creación particular. Dios crea cada alma individual de tal modo que la relación entre ella y su cuerpo no es para el cristiano, como había sido en el planteamiento platónico, accidental: la noción de alma tiene para el cristiano, pues, un carácter personal. Excediendo asimismo el planteamiento aristotélico, la noción cristiana de alma se sitúa más allá de todo proceso meramente biológico o intelectivo, apareciendo antes bien vinculada con el plano de lo espiritual: la vida espiritual del alma cristiana poco o nada tiene que ver con los procesos vitales a cuya base situara Aristóteles su noción de alma.

Tras la etapa medieval, que acabamos de sobrevolar sin entrar en ningún tipo de detalle, la filosofía moderna comienza a cobrar conciencia del problema de la conciencia. Han sido muchos los que en Descartes han encontrado la primera formulación moderna del problema mente-cuerpo (Rozemond, 2006: 48). Y es que, a pesar de que fuera el francés un pensador religioso, su ontología excede el ámbito de la especulación de carácter teológico: en ella no se trata de las relaciones entre el alma, interpretada en términos espirituales, y la materia en un universo teológico, sino de las relaciones entre la mente, descrita como una entidad dotada de capacidades intelectuales, y un cuerpo incardinado en el mundo gobernado por las leyes mecánicas matematizadas que la nueva física empezaba a iluminar. Descartes se aleja asimismo del planteamiento clásico de la cuestión al desvincular completamente el cuerpo y la mente: al contrario de lo que sucedía en la tradición griega, la mente ya no es en Descartes principio de vida y movimiento. La mente se caracteriza por el pensamiento, netamente inextenso, mientras el

[6] Como atinadamente señala Tarnas (1991: 55 del original, 86 de la traducción), es más que probable que esta clase de enredos hermenéuticos se deban en el caso de Aristóteles a los avatares de nuestra recepción de su obra.

cuerpo, extenso, se halla sometido a las leyes mecánicas que rigen el movimiento y de ningún modo podría dar lugar a pensamiento alguno: las almas inferiores de la psicología aristotélica, ésas que animaban el cuerpo, son sustituidas en Descartes por leyes mecánicas. La doctrina cartesiana ha sido tradicionalmente denominada *dualismo sustancialista*, dado que en ella la mente y el cuerpo son definidos como dos sustancias totalmente heterogéneas. Tendríamos, pues, por una parte, *res extensa*, la sustancia corporal, cuyo atributo característico es la extensión –una sustancia material y sometida a leyes mecánicas–, y, por otra, *res cogitans*, la sustancia mental, cuyo atributo característico es el pensamiento –una sustancia inextensa, inmaterial, concebida como puro pensamiento libre del imperio de las leyes mecánicas–. Descartes se anticipa de este modo a la formulación de uno de los núcleos del contemporáneo problema de la conciencia: difiriendo en la medida en que parecen diferir las propiedades de los objetos materiales y las de los estados mentales conscientes –en el marco cartesiano, esencialmente porque los estados mentales son distintivamente no espaciales, por contraposición a los objetos materiales, ellos sí, extensos–, ¿cómo esbozar un marco teórico en el que ambos encajen, un marco teórico que permita diluir su aparente heterogeneidad y avanzar hacia la comprensión de su interrelación?.

Con la señalada heterogeneidad entre la sustancia corporal y la mental postulada por Descartes hace acto de presencia en el ámbito de la reflexión filosófica acerca de la mente el moderno problema mente-cuerpo, dado que con ella surge el problema de elaborar un esquema explicativo capaz de dar cuenta del modo en que mente y cuerpo se relacionan, el cual es a menudo traducido a un determinado léxico contemporáneo que lo presenta como el de elucidar el modo en que la actividad neurofisiológica deviene en actividad mental. No obstante, no son éstos los términos en que a Descartes se le presentara el problema: él tenía que explicar, por su parte, cómo podían interactuar la dos sustancias completamente heterogéneas que había postulado. ¿De qué modo llegan a la mente esos procesos mecánicos que parten del cuerpo y dan, justamente, cuerpo a la percepción? ¿De qué modo las intenciones de la mente se convierten en movimientos del cuerpo? Es con este tipo de preguntas que surge el moderno problema mente-cuerpo, del que, como señalábamos, el contemporáneo problema de la conciencia puede ser considerado una refinación. Así, las posturas tradicionales en el problema de la conciencia surgen del intento de solucionar los dilemas ínsitos en la ontología cartesiana. La solución de Descartes, como es sabido, apela a la existencia de un centro de interacción entre las sustancias pensante y extensa que se encontraría en la glándula pineal,

pues consideraba que la misma era especial dado que se halla entre ambos hemisferios cerebrales, no apareciendo, a diferencia de la mayoría de las estructuras encefálicas, duplicada. Además, suponía erróneamente que se trataba de un centro nervioso que sólo podía encontrarse en el encéfalo humano. Estas peregrinas consideraciones le llevaron a aseverar, pues, que la necesaria comunicación entre el mundo inmaterial de lo mental y el mundo extenso de lo corpóreo tenía lugar, de algún misterioso modo, a través de la glándula pineal.[7] Esta solución cartesiana se ha denominado tradicionalmente *dualismo interaccionista*. La principal entre las debilidades del dualismo interaccionista, una debilidad que ni Descartes ni ningún otro teórico posterior ha alcanzado a superar, se halla en el modo en que dicha propuesta, como sugiriéramos ya en el exordio, viola la ley de conservación de energía. Descartes trató de dar respuesta a este problema mediante un ingenioso malabarismo erístico que, para desgracia del dualista, no ha resistido muy bien el paso del tiempo.

> Descartes a reconnu, que les âmes ne peuvent point donner de la force aux corps, parce qu'il y a toujours la même quantité de force dans la matière. Cependant il a cru que l'âme pouvait changer la direction des corps. Mais c'est parce qu'on n'a point su de son temps la loi de la nature, qui porte encore la conservation de la même direction totale dans la matière (Leibniz, 1714: §80).[8]

En otras palabras, si Descartes hubiera dispuesto de la noción de momento, no habría intentado defender su interaccionismo de la acusación de violar la señalada ley de conservación mediante una estrategia consistente en pretender que la mente puede cambiar la dirección de un corpúsculo, pero no su cantidad de movimiento. En otras palabras, Descartes fundaba su defensa del dualismo interaccionista en este punto en la curiosa opinión de que –en sus términos (vid., Descartes, 1637)– el movimiento es algo diferente de la determinación de la dirección y, así, la fuerza es necesaria para mover un cuerpo, pero no para determinar la dirección en la que el mismo ha de moverse.

Otros filósofos racionalistas trataron de arrostrar el desafío teórico cartesiano elaborando dos formas diferentes de *dualismo sustancialista* y una de *monismo*. Por lo que a este último respecta, Baruch de Spinoza, partiendo de una reformulación de la metafí-

[7] Hoy sabemos que la glándula epitalámica de secreción interna que Descartes escogiera como estación de relevo mente-cuerpo es, irónicamente, la encargada de la secreción de melatonina, una hormona implicada en el tránsito de la vigilia al sueño, y que no cumple ninguna función relevante en el desempeño de actividad cognitiva alguna. Por otra parte, tanto la glándula como la hormona es posible que deparen aún sorpresas (vid., v. g., Vincent, 2007: 64 de la traducción).

[8] Sería justamente el abandono de la idea de que el alma puede variar la dirección (pero no la cantidad) del movimiento uno de los principales estímulos para el desarrollo del paralelismo leibniziano.

sica cartesiana, sostuvo que la extensión y el pensamiento no deben considerarse, como hiciera Descartes, dos sustancias diferentes, sino dos aspectos o atributos de una única sustancia, que Spinoza identificaba con Dios o la naturaleza (*Deus sive natura*). Es, para Spinoza, una y la misma cosa la que piensa y es extensa. La postura del holandés ha sido en este sentido calificada de *monismo neutral*, dado que en la misma se nos habla de diversas facetas de una y la misma sustancia: lo mental y lo material serían desde esta perspectiva una y la misma cosa vista desde deferentes ángulos. De este modo conseguía Spinoza ofrecer respuesta al hecho de que dos esferas aparentemente tan diversas como las de lo mental y lo corporal se hallaran coordinadas: la correspondencia entre ambos mundos se debería, en definitiva, a que de hecho no hay más que uno solo.

Por lo que toca a las dos variaciones sobre la línea melódica del dualismo sustancialista que surgieran en respuesta a las dificultades que el planteamiento cartesiano había puesto sobre la mesa, el *ocasionalismo* de Malebranche y la *armonía preestablecida* de Leibniz fueron las dos alternativas al monismo que surgieran dentro del seno de la filosofía racionalista.[9] La solución de Leibniz a los problemas que Descartes abordara mediante su interaccionismo difiere ostensiblemente de éste: en lugar de postular una interacción entre dos esferas ontológicas tan diversas, el alemán sugiere que Dios las tendría bien dispuestas desde el propio momento de la creación para que sus procesos coincidieran marchando paralelamente a la par, como dos relojes diferentes pero perfectamente sincronizados. Malebranche, por su parte, involucra nuevamente a un ser todopoderoso para resolver el problema de la relación entre la mente y el cuerpo, pero el Dios de Malebranche se ve obligado a realizar un trabajo un poco más pesado que aquél que Leibniz pidiera al suyo: en lugar de tener que diseñar el espectáculo de la perfecta coordinación entre mente y cuerpo en el momento de la creación, el Dios de Malebranche tiene que intervenir para propiciar esa coordinación todas y cada una de las veces

[9] En realidad, hablar de la postura leibniziana como una alternativa al monismo resulta un tanto comprometido, dado que Leibniz entendía que la concepción cartesiana de la sustancia extensa, interpretada en términos mecanicistas como algo inerte o inactivo, como algo que no tiene en sí el principio de su movimiento, no concuerda con la autosuficiencia otorgada por definición a la noción de sustancia. Esto traería consigo su rechazo del dualismo sustancialista cartesiano. En su ontología el universo se halla constituido por infinitas sustancias individuales, cada una de las cuales posee el principio de sus propias modificaciones como algo inherente. De modo que Leibniz viene a llenar el abismo que Descartes había situado entre pensamiento y extensión al integrar ese principio activo –mediante el recurso a la noción aristotélica de forma sustancial– con la "simple materia extensa" en una unidad en la que ninguno de ambos elementos actúa independientemente, sino que el activo es causa del movimiento del pasivo, del cual depende asimismo para desarrollar su acción, ya que cada átomo sustancial (y no mera o puramente extenso) es una unión indisoluble de ambos, de lo corpóreo y la mente. Así, según Leibniz, la sustancia corpórea contiene, por una parte, un principio activo –la forma, el alma– y, por otra, un principio pasivo –la materia extensa–, y ambos principios constituyen un *unum per se*.

que nos rascamos porque nos pica y todas y cada una de las veces que nos duele porque nos rascamos demasiado; esto es, el Dios de Malebranche apenas tiene tiempo para hacer otra cosa más allá de producir movimientos corporales correspondientes a actos mentales y viceversa. Las propuestas de Leibniz y Malebranche pueden ser englobadas bajo la etiqueta «paralelismo». Según el paralelismo, en síntesis, nuestras ganas conscientes de salir a la calle y el desplazamiento de nuestro cuerpo hasta la misma son dos fenómenos totalmente desconectados y su interacción no sería más que una ilusión causada por el hecho de que Dios se habría tomado la molestia de sincronizar con precisión los fenómenos pertenecientes a las esferas de lo mental y lo corpóreo.

El problema de abordar las relaciones entre ese mundo que denominamos para abreviar "material" y ese otro que por idénticos motivos denominamos "mental" resulta complejo al punto que las mentes más preclaras del racionalismo no supieron sino inmiscuir en su elucidación potencias sobrenaturales lo suficientemente amables como para ahorrarnos la necesidad de arrojar por nuestros medios claridad sobre el mismo. La última de las soluciones clásicas al problema mente-cuerpo, gestada en el seno de la tradición empirista, recurre idénticamente a un ser todopoderoso tan solícito como para resolver en nuestro lugar el enigma psicofísico. Como indicábamos, desde la perspectiva ofrecida por el paralelismo nuestra impresión de que los fenómenos mentales y los corporales se hallan causalmente vinculados puede ser tenida por ilusoria. En el seno de la filosofía empirista surge, como apuntábamos, la última entre las posiciones clásicas en el problema de la conciencia, el *idealismo subjetivo* de Berkeley, según el cual la ilusión es aún mayor: nuestra impresión de que el mundo material, efectivamente, existe es ahora lo ilusorio. El problema de coordinar las esferas ontológicas postuladas por Descartes desaparece desde el momento en que una de ellas es concebida como un sueño de la otra: sólo la mente y sus contenidos existen mientras los objetos que percibimos tienen el mismo estatus ontológico que los objetos que aparecen en nuestros sueños. La realidad última de dichos objetos es la de una colección de ideas infundidas en la mente de los seres humanos por la de Dios.

En el siglo posterior a la muerte de Berkeley la biología evolucionista de Darwin da a Spencer la clave para cimentar una nueva concepción del problema mente-cuerpo: lo mental surge en la filogénesis de la mano de la evolución del sistema nervioso. No es habitual encontrar mencionada en manuales de filosofía de la mente esta postura naturalista entre las soluciones clásicas al problema mente-cuerpo. Nosotros hemos optado por cerrar este apartado aludiendo brevemente a ella, por una parte, por el creciente e inne-

gablemente justificado impacto de este germen decimonónico en la filosofía de la mente y las ciencias cognitivas y, por otra, porque entendemos que, efectivamente, "with the publication of the *Origin of Species* everything changed. Nothing would quite be the same again. It's important to realise that Darwin's most famous work not only changed our ideas about how species came about, but also radically reformed our view of the natural world and our place within it" (Workman, 2014: 16).

## 2._Planteamiento actual

Aunque la división pueda resultar artificial, cabe distinguir un problema explicativo y un problema ontológico de la conciencia. Con arreglo a esta distinción, delinearemos en este capítulo un esquema de los planteamientos actuales que tratan de especificar el lugar que la conciencia ha de ocupar en nuestra concepción del mundo (perspectivas de corte ontológico) y, en el siguiente, haremos lo mismo con los que tratan de ofrecer una explicación de la conciencia por lo que toca a su origen y mecanismos (perspectivas explicativas). En la práctica, en las propuestas de los diferentes filósofos y científicos implicados en el debate contemporáneo, aspectos de ambas formas de abordar el problema de la conciencia aparecen a menudo inevitablemente entreverados, de ahí que, como sugeríamos, constituya esta distinción un artefacto expositivo antes que una descripción enteramente ajustada a una limpia y tajante división teórica realmente existente.

La principal segmentación que cabe trazar entre los diferentes intentos destinados a delimitar el lugar que la conciencia ha de ocupar en nuestro esquema de la realidad sería equivalente a la tradicional distinción entre monismos y dualismos. A pesar de que cabe distinguir entre monismos materialistas y monismos idealistas (como el idealismo subjetivo de Berkeley), dado que estos últimos no cuentan actualmente con partidarios, presentaremos de forma sucinta las posibilidades dualistas aún vigentes (incluyendo en este grupo las perspectivas neutrales o del doble aspecto) para pasar a ocuparnos a renglón seguido de los diferentes monismos fisicalistas contemporáneos.

## 2.1._Dualismos

*Dualismos sustancialistas*

El *dualismo sustancialista interaccionista* de corte cartesiano prácticamente ha desaparecido del panorama contemporáneo. No obstante, siguen encontrándose entre los filósofos contemporáneos algunos defensores de dicha postura (vid. Foster, 1989; 1991; Swinburne, 1986; 2013), pero la verdad es que sus propuestas no tienen demasiada repercusión dentro del actual debate, limitándose a jugar el papel de postura cuyas falencias todo el mundo se afana en evitar. También entre los científicos han sido minoría los que en las últimas décadas han defendido este tipo de dualismo. El eminente neurofisiólogo australiano John C. Eccles encarnó una sonada excepción a esta norma. En una línea que prolongara la trazada por Charles Sherrington –uno de los más destacados neurocientíficos de la generación anterior–, Eccles sostuvo que si bien la conciencia debió originarse en el proceso de la evolución biológica, la misma ha de ser concebida como algo inmaterial aunque capaz de intervenir en el mundo material a través del influjo que ejercería en nuestra conducta mediante su influencia sobre el cerebro. El dualismo interaccionista de Eccles trataba de eludir el epifenomenalismo postulando 1) que la conciencia integra la información que puede encontrarse en los diferentes módulos del neocórtex produciendo una escena unificada, y 2) que, como efecto de la decisión consciente, la entelequia inmaterial que Eccles denominara "mente autoconsciente" es capaz de activar los circuitos neuronales apropiados para la producción de las eferencias motoras necesarias para llevar a cabo la conducta correlativa a la señalada decisión consciente. Nos encontramos de este modo ante un Descartes del siglo XX[10] que desde la década de los setenta (vid. Eccles, 1974) vino perfilando una propuesta interaccionista en la que la glándula pineal cartesiana aparecía disfrazada de conjuntos de neuronas dispuestas modularmente en el hemisferio dominante (Eccles, 1980; 1981), unos módulos que, según este esquema interaccionista –presentado en términos asemánticos (Lamote de Grignon, 2005: 446) y tan extravagantes como *psicones* afectando a *dendrones* (Eccles, 1990)–, están abiertos a las influencias de la mente autoconsciente.

[10] En Eccles (1953) el australiano elaboró, de hecho, un planteamiento interaccionista idéntico en esencia al de Descartes. Posteriormente, el desarrollo de su propia disciplina le obligaría a abandonar el supuesto de un centro cerebral específico como *locus* de la interacción mente-cerebro.

En sus últimos escritos al respecto, Eccles perseveraría en su dualismo interaccionista perfilando, entre otros dispositivos retóricos, un modelo del control motor consciente basado en la física cuántica (vid. Beck & Eccles, 1992; 2003). En un intento por salvar el dualismo interaccionista de la principal objeción de la que ha sido objeto –la de que viola las leyes de conservación de la física, como indicáramos –, Eccles acabaría por proponer que la conciencia puede afectar a procesos cuánticos acaecidos en la membrana terminal de la neurona presináptica incrementando la probabilidad de que en ésta las vesículas sinápticas liberen su contenido al espacio sináptico. Influyendo así sobre procesos probabilísticos a nivel sináptico en amplias regiones del sistema nervioso, la conciencia inmaterial que Eccles persiguió toda su vida podría dar cuenta de los movimientos voluntarios y, así, de la interacción entre la mente inmaterial y el cuerpo y su conducta, pero lo cierto es que un modelo tan difícil de operacionalizar podría dar cuenta de prácticamente cualquier cosa.

Recientemente el físico de partículas Henry Stapp ha defendido también un planteamiento dualista interaccionista basado asimismo en la física cuántica, pero partiendo de supuestos diversos de los de Eccles y Beck. Su modelo no necesita, como el de Eccles, que la conciencia sesgue las probabilidades de diversos estados cuánticos, y de hecho considera que tal estrategia argumentativa entra en contradicción con el modelo estándar de la mecánica cuántica. La propuesta de Stapp se halla antes bien destinada a ofrecer una reinterpretación del tradicional problema de la delimitación del observador y el sistema observado en física cuántica. Partiendo de la concepción de Neils Bohr, la física cuántica no describe un mundo físico caracterizado por su autonomía respecto del sujeto descriptor, sino las observaciones hechas por observadores conscientes en virtud de determinados protocolos de medida. No obstante, en la concepción de Bohr la frontera entre observador y sistema observado fue trazada de tal forma que los procesos que formaban parte del observador incluían su mente, su cerebro, su cuerpo y los instrumentos por él manejados. ¿Cuál de estos elementos es el responsable de que el sistema observado colapse ofreciendo al observador una observación y no otra? La solución podría consistir en trazar en alguno de los puntos señalados la frontera entre sistema observado y observador, pero von Neumann demostró que nada en el aparato teórico de la mecánica cuántica permite excluir ninguno de los elementos físicos señalados de la descripción cuántica del sistema observado. La única frontera que parece quedarnos de este modo es la que conformaría el interfaz entre la experiencia consciente que las formalizaciones cuánticas describen y el cerebro del observador. Stapp desarrolla esta idea para presen-

tar a las decisiones conscientes como agentes del colapso. En el cerebro del propio observador se encontrarían las virtualidades cuánticas a las que aquellas decisiones afectan. De este modo, la conciencia escoge las preguntas a realizar, que por mediación de su efecto en el cerebro del observador interactúan con las posibilidades en curso especificadas por la mecánica del sistema sobre el que dichas preguntas versan, incluyendo el propio cerebro (vid. Stapp, 2007; 2007/2011). Como puede observarse, en el caso de Stapp, no nos encontramos con un intento de desentrañar la naturaleza de la conciencia o hacer comprensible la existencia de la misma en y por sí misma, sino por el modo en que ella ha venido viéndose complicada con los desarrollos de la física de partículas no clásica.

La otra forma de dualismo metafísico, el *paralelismo psicofísico*, no goza de ningún predicamento en la actualidad.

*Dualismos de propiedades*

El *dualismo de propiedades*, por el contrario, es actualmente una postura muy ramificada y con gran cantidad tanto de partidarios como de detractores. Lo primero que debemos señalar es que el dualismo de propiedades es ya, a diferencia de los dualismos comentados previamente, un monismo metafísico, pues sostiene que, aunque las propiedades fenoménicas de los estados mentales conscientes no sean reductibles a propiedades físicas, las mismas han de ser instanciadas por aquello mismo en lo que se realizan las propiedades físicas. La de «dualismo de propiedades» es una etiqueta un tanto general, por cuanto comprende subtipos tales como el *monismo neutral* de corte russelliano (desde el punto de vista del cual tanto las propiedades conscientes de los estados mentales como las propiedades físicas son derivadas o dependientes de un constituyente más básico de la realidad; el Strawson de los noventa contaría como principal defensor contemporáneo de esta postura), el *monismo anómalo* de Davidson (que sostiene que cada suceso mental particular es idéntico a un suceso físico particular, pero que no existen estrictas leyes naturales que conecten la esfera de lo mental definida en términos mentales con la de lo físico), el *naturalismo biológico* de Searle (empeñado en no ser considerado dualista de propiedades)[11], el *epifenomenalismo* de Jackson (que, basándose en

[11] No nos detenemos aquí en una breve descripción de esta postura porque en el primer capítulo de la tercera parte ofreceremos un detallado análisis de la misma.

argumentos epistemológicos acerca de nuestra capacidad para comprender aspectos cualitativos de la experiencia consciente partiendo de datos de tercera persona, propone una ontología en la que dichos aspectos se presentan como irreductibles a cualesquiera propiedades físicas y, además, como causalmente inertes), el *dualismo de propiedades emergentista* de Hasker (según el cual los estados mentales conscientes tienen su origen en la organización de elementos puramente físicos, pero con su emergencia surge algo por encima y más allá del substrato físico de la misma, y además, dicha emergencia ni es predecible a priori ni es lo emergido explicable en términos de leyes naturales referidas a su substrato físico),[12] el reciente *panpsiquismo* de Strawson (de acuerdo con el cual todos los niveles de la realidad, desde los más básicos, poseen propiedades mentales o proto-mentales diversas de la propiedades físicas que asimismo poseen), el *panprotopsiquismo* de Chalmers (en el marco del cual esas propiedades mentales fundamentales dentro del esquema de la realidad se presentan como propiedades protopsíquicas de los constituyentes mínimos de la realidad que, en determinadas circunstancias, pueden ocasionar verdaderos estados mentales conscientes) y su *dualismo naturalista* (denominado también *dualismo de propiedades fundamental*, por cuanto entiende que las propiedades mentales fenoménicamente conscientes forman parte de los constituyentes últimos, fundamentales o básicos de la realidad, que dichas propiedades mentales pueden interactuar causal y legaliformemente con otras propiedades fundamentales, pero que su existencia es ontológicamente independiente de cualesquiera otras propiedades básicas: no derivan ni dependen de ellas y no pueden reducirse a o ser explicadas por las mismas).[13]

En último término, talvez la única la característica compartida por todas las posturas que hemos clasificado como dualistas de propiedades sea la defensa de la autonomía de lo mental respecto de lo físico.

[12] Vid. Hasker (1999).

[13] Hierro-Pescador (2005: 187), así como el propio Chalmers (1996: 168), han hablado explícitamente de la postura de éste como un dualismo de propiedades. Por el contrario, muchos autores eluden la etiqueta o la rechazan a pesar de que al resto de los especialistas les resulte meridianamente claro que la postura que defienden cae dentro del fardo marchamado con la señalada signatura.

## 2.2._Monismos fisicalistas

*Las teorías de la identidad*

Cabe incluir en primer lugar bajo el epígrafe "monismos fisicalistas", siguiendo el criterio cronológico, a las denominadas teorías de la identidad. Con todo, es necesario matizar el sentido en que resulta pertinente incluir a las teorías de la identidad bajo el epígrafe "fisicalismos" en un apartado dedicado a las perspectivas ontológicas en el debate actual en torno al problema de la conciencia, dado que dicho marco teórico fue elaborado en la década de los cincuenta, un momento previo a la conformación del contexto intelectual en el que, desde finales de los ochenta, viene discutiéndose el problema de la conciencia. En el momento en que se gestaran las teorías de la identidad, pues, el debate estaba orientado a la cuestión más general de la ontología de lo mental. Nos encontramos, en resumidas cuentas, en los años de formación de la disciplina que hoy denominamos filosofía de la mente y, desde luego, no cabe esperar del debate sostenido en aquel contexto filosófico y aquel entonces respuestas específicas a muchos de los extremos que hoy se discuten en cada una de las ramas que surgieran de aquella entonces incipiente disciplina. No obstante, las teorías de la identidad abordan algunas de las cuestiones actualmente en liza en los *Consciousness Studies* y, de este modo, no han dejado de influir directamente en la configuración del debate contemporáneo acerca del problema de la conciencia. Así, por una parte, cabe entender que las propuestas funcionalistas, que tan decisivamente han contribuido a configurar dicho debate, surgieran en cierta medida en respuesta a las teorías de la identidad y en cierta medida como extensión de una de las versiones en que éstas se presentaron y, por otra, dichas teorías pueden ser consideradas como el referente fundamental de la mayor parte de las variantes del fisicalismo que posteriormente trataron de distanciarse del funcionalismo. Además, y a pesar del general abandono de las teorías de la identidad a partir de la década de los sesenta, algunos investigadores en el área de los *Consciousness Studies* han vuelto recientemente a echar mano de sus supuestos en su tratamiento de algunos de los puntos centrales de la agenda de la disciplina (vid., v. g., Hill & McLaughlin, 1998; Papineau, 1995; 2002).

La tesis central de las teorías de la identidad puede resumirse como sigue: los estados y procesos que denominamos mentales son estados y procesos del sistema nervioso. No se trata de que entre ambos se de una correlación o correspondencia, sino que

aquéllos se identifican con éstos. En las antípodas del dualismo sustancialista cartesiano, el mundo mental y el mundo físico dejan de constituir dos categorías ontológicas excluyentes y, en lugar de ello, los fenómenos mentales nos son presentados por los teóricos de la identidad como una serie de estados y procesos acaecidos en determinados segmentos de determinados sistemas físicos, en concreto, en el sistema nervioso de al menos un animal: *Homo sapiens*. El problema que presentáramos más arriba como germen moderno del debate contemporáneo en torno al problema de la conciencia, el de dar cuenta del modo en que el mundo físico y el mental se hallan relacionados, sencillamente desaparece desde el momento en que lo mental se identifica con una parte del mundo físico. No puede hablarse de interacción entre dos cosas cuando contamos sólo con una.

Insatisfechos con el modo de eludir el dualismo que ensayaran los conductistas analíticos, los teóricos de la identidad pergeñarían un conato similar: en lugar de identificar los fenómenos mentales con disposiciones a la conducta, lo hicieron con estados del sistema nervioso. Los primeros filósofos en explorar esta vía fueron Ullin T. Place y Hebert Feigl, en sendos artículos publicados entre 1956 y 1958: "Is consciousness a brain process?", del primero, y "The 'mental' and the 'physical'", del segundo. La publicación del primero siguió a una serie de reuniones y discusiones entre Place, John Jamieson Carswell Smart, Charles Burton Martin y Douglas Aidan Trist Gasking que tuvieron lugar en 1954 en la Universidad de Adelaida. En ellas, Place logró que Smart abandonara su adhesión al conductismo de Ryle. El fruto más temprano de este cambio de orientación de las perspectivas de Smart en filosofía de la mente sería el clásico "Sensations and brain processes", de 1959. Por lo que toca al artículo de Feigl, y tal y como el propio autor ha sido el primero en señalar (vid. Feigl, 1981: 288), el contenido del mismo puede entenderse como un desarrollo de la teoría del doble lenguaje que Moritz Schlick defendiera en su obra temprana *General Theory of Knowledge* –publicada en 1918, cuatro años antes de la conformación del Círculo de Viena–, una propuesta según la cual la diferencia entre lo mental y lo físico no es una diferencia entre dos ámbitos de la realidad, entre dos esferas ontológicas, sino entre dos sistemas conceptuales. Estos tres artículos pueden ser leídos como documentos fundacionales de las teorías de la identidad, aunque no pocos mantendrían que cabría hacerlo antes bien como sus certificados de nacimiento y defunción. Nos aproximaremos brevemente a continuación a la tradición que estos artículos inauguran y comprobaremos al hacerlo la adecuación del plural que venimos empleando al utilizar la etiqueta usualmente asignada a los planteamientos de estos filósofos: *teorías de la identidad*.

La propuesta de Place es interesante en el contexto de los *Consciousness Studies* dado que deja de lado las posibilidades que el conductismo lógico o analítico pudiera tener de cara a dar cuenta de términos cognitivos como «comprender», «recordar» o «creer» y da por supuesto que una aproximación conductista a términos vinculados con la experiencia consciente (con eso que hoy denominaríamos *qualia*) sería en cualquier caso insuficiente. En este contexto, Place planteó su pionera teoría de la identidad en relación con dicha clase de términos mentalistas de forma totalmente explícita: "«Consciousness is a process in the brain», in my view is neither self-contradictory nor self-evident; it is a reasonable scientific hypothesis in the way that the statement «lightning is a motion of electric charges» is a reasonable scientific hypothesis" (Place, 1956: 45). Así, al presentar su tesis como una hipótesis científica, Place entiende que la misma no puede refutarse en base a meros argumentos de carácter lógico. Además, desde su punto de vista, las objeciones de carácter lógico que cabría hacer a la identidad constitutiva contenida en el primer enunciado son equivalentes a las que cabría hacer a la expresada en el segundo. En este sentido, no constituiría amenaza alguna a su hipótesis señalar que «sensación» y «actividad neurofisiológica» significan cosas diferentes, dado que también «rayo» y «movimiento de cargas eléctricas» tienen significados diferentes, pues no sabemos que un rayo es un movimiento de cargas eléctricas sino por la investigación y la teoría científica, cosas de las que perfectamente pueden prescindir hablantes competentes al usar con perfecto sentido la expresión «rayo». Así, «sensación» y «procesos neurofisiológicos» pueden vincularse mediante la cópula «es» en tanto descripciones de una y la misma cosa, a pesar de que impliquen modos diversos de acceso epistémico a la misma. Según Place, en definitiva, el paralelo a trazar implicaría al enunciado de identidad «la conciencia es un proceso neurofisiológico» y cualquier otro en el que sea predicada una identidad entre algo cuya ocurrencia sea verificada mediante procesos ordinarios de observación, por una parte, y algo cuya ocurrencia sea establecida mediante alguna clase especial de procedimiento científico, por la otra (Place, 1956: 47). Para acabar, Place añade a su argumentación la noción de "falacia fenomenológica", que describe como la idea de que cuando un sujeto habla de sus experiencias conscientes lo que hace es describir las propiedades reales de objetos y eventos proyectados en una especie de pantalla interna de cine. La crítica de esta falacia se ha prolongado hasta nuestros días, siendo, como veremos, desarrollada en detalle por Daniel Dennett.

Por su parte, Feigl y Smart alcanzaron conclusiones análogas haciendo uso de la distinción, de larga tradición en filosofía del lenguaje, entre referencia y sentido al ar-

gumentar que a pesar de que «sensación» y «actividad neurofisiológica» puedan tener diferentes sentidos, no hay motivos por los que deban tener referentes distintos. En cualquier caso, ni los planteamientos de éstos ni los de aquél sugieren, como los de algunos teóricos posteriores (vid. Kripke, 1980), que los enunciados de identidad hayan de implicar identidades necesarias.

Un paso más allá dentro de la tradición de las teorías de la identidad sería dado por David Lewis, Brian Medlin y David Malet Armstrong en la década de los sesenta al subrayar el carácter causal de lo mental, apareciendo así como antecedentes del funcionalismo en filosofía de la mente. En este sentido, el primero llegaría a afirmar (Lewis, 1966) que la característica definitoria de toda experiencia es su rol causal, una afirmación que desconcertará al lector de filosofía de la mente contemporáneo, acostumbrado a zombies, espectros invertidos y, en definitiva, a la opinión de que precisamente son los *qualia* el único aspecto de lo mental que escapa al intento de analizar lo mental en términos funcionalistas, en términos de roles causales. No obstante, a diferencia de los funcionalistas posteriores, Lewis no ofrece en el texto citado una caracterización abstracta de su noción de rol causal, sino que la vincula directamente con estados concretos de sistemas físicos, los cuales, en su análisis, constituirían las propias experiencias conscientes.

La aproximación a planteamientos funcionalistas por parte de teóricos de la identidad puede ser iluminada mediante la distinción entre teorías de la identidad de tipos (type-type identity theories) y teorías de la identidad de casos (token-token identity theories). Según las teorías de la identidad de tipos, cuando dos personas comparten el mismo estado mental, comparten el mismo estado del sistema nervioso: nos hallamos ante dos casos del mismo tipo de estado mental y dos casos del mismo tipo de estado del sistema nervioso. Desde esta perspectiva, partiendo del hecho de que dos personas se hallan en el mismo estado mental, podemos concluir que sus sistemas nerviosos se hallan en ese momento en el mismo estado, y viceversa. Por su parte, según las teorías de la identidad de casos, dos personas que comparten el mismo estado mental, pueden no compartir el mismo tipo de estado de sus sistemas nerviosos: nos hallamos ante dos casos del mismo tipo de estado mental, pero no necesariamente ante dos casos del mismo tipo de estado del sistema nervioso. Desde esta perspectiva, partiendo del hecho de que dos personas se hallan en el mismo estado mental, no podemos concluir que sus sistemas nerviosos se hallen en ese momento en el mismo estado y, sin embargo, partiendo del hecho de que el sistema nervioso de dos personas se hallan en el mismo esta-

do, sí podríamos concluir que ambas comparten en ese momento el mismo estado mental. La línea que de las teorías de la identidad partiera hacia la elaboración del posterior marco teórico funcionalista se desarrollaría en congruencia con el núcleo teórico de las teorías de la identidad de casos.

*El funcionalismo*

El funcionalismo, a diferencia de las teorías de la identidad, no concibe los estados mentales como idénticos a estados del sistema nervioso, sino como realizados en el soporte material del sistema nervioso. Este amplio marco teórico goza actualmente de un gran predicamento en filosofía de la mente y es frecuente encontrarlo definido como el substrato filosófico de las ciencias cognitivas (Moya, 2004: 105), pues, como Jaegwon Kim ha señalado, el funcionalismo parecía hecho a medida para esta nueva ciencia de lo mental (Kim, 1988a: 48) –aunque en verdad sólo pueda esto afirmase hoy entre comillas dada la actual pujanza de paradigmas en ciencias cognitivas que exceden el tradicional representacionalismo que, en efecto, casaba a la perfección con el funcionalismo–. Hay que matizar antes de entrar en detalles que, al igual que las teorías de la identidad, el funcionalismo surge como un marco teórico abarcador, es decir, destinado a ofrecer una conceptualización general de lo mental y no sólo de la mente consciente. Con todo, su influencia en el planteamiento contemporáneo del problema de la conciencia puede apreciarse ya en el hecho de que la inmensa mayoría de los argumentos elaborados por filósofos interesados en demostrar que la experiencia consciente es inabordable dentro de un marco teórico fisicalista han venido tratando de desafiar el entramado teórico funcionalista. Este hecho obedece a que el funcionalismo, tras comenzar a prevalecer en la filosofía analítica de la mente a principios de los sesenta sobre el conductismo lógico y las teorías de la identidad, y, asimismo, tras sucesivos y significativos apuntalamientos y modificaciones, ha venido siendo la propuesta teórica más exitosa en el mercado de la filosofía de la mente, la ortodoxia, cabría decir, en filosofía de la mente.

La nota característica del funcionalismo es su tratamiento de las propiedades y estados mentales en términos de propiedades funcionales, es decir, en términos del rol causal que desempeñan. Los estados mentales son, desde esta perspectiva, estados caracterizados por su posición en una cadena causal, en la cual intervienen como causas y como efectos. Un estado mental es analizado en este marco teórico como efecto de cier-

tos estímulos externos o de otros estados mentales y, al mismo tiempo, como causa de una determinada conducta o de otros estados mentales (Hierro-Pescador, 2005: 93). La tesis central del funcionalismo sostiene, pues, que un estado mental es el estado que es no en virtud de que el sujeto del mismo se halle en un determinado estado neurofisiológico, como en las teorías de la identidad (que implicaban la reducción de la psicología y su vocabulario a la neurofisiología y el suyo), sino en virtud de una determinada cadena de relaciones causales entre entradas perceptivas (inputs), otros estados mentales y respuestas conductuales (outputs) en la que el estado mental en concreto juega un papel definido. Es el papel causal que un estado mental desempeña en la economía de un organismo o sistema el cabo del que tirar a la hora de individuarlo e identificarlo.

Lo mental deja con el funcionalismo de identificarse con disposiciones a la conducta o estados del sistema nervioso central, como sucedía dentro del marco teórico del conductismo lógico y las teorías de la identidad, y pasa a interpretarse no directamente en relación con estados, eventos o propiedades físicas, sino en relación con lo que esos estados, sucesos y propiedades hacen. "«Handsome is as handsome does» (…) la materia importa sólo por lo que es capaz de hacer" (Dennett, 2005: 16 del original, 32 de la traducción). No obstante, el funcionalismo no se aparta de una perspectiva fisicalista, pues, aunque dos estados puedan resultar equivalentes funcionalmente y diversos atendiendo a su implementación física, se considera que no puede darse diferencia funcional sin diferencia física, es decir, se postula la dependencia de lo funcional respecto del nivel físico de implementación, lo cual viene a significar que no cabe concebir identidad en lo material y diversidad en lo funcional (mental), esto es, no se considera posible que varíen aspectos funcionales (mentales) permaneciendo idéntica la base material. Esto es lo que se ha venido llamando en filosofía de la mente –desde Davidson– relación de superveniencia, de la cual se puede partir hacia la comprensión de la polémica tesis de la realizabilidad múltiple, según la cual idénticos estados funcionales (mentales) pueden ser realizados sobre la base de diversos substratos materiales (Putnam, 1967; Fodor, 1974), siendo así que un conocimiento detallado de cualquiera que sea el substrato material de la realización (v. g., el sistema nervioso central mamífero) será en cualquier caso insuficiente –y hasta irrelevante– de cara a comprender la naturaleza y funcionamiento de lo mental. Una consecuencia antiintuitiva de esta tesis sería que sistemas radicalmente distintos de nosotros podrían tener estados mentales como los nuestros siempre y cuando porten estados que desempeñen los papeles causales desempeñados por nuestros estados mentales, es decir, siempre que estén dotados de la organiza-

ción funcional adecuada, con independencia de su constitución física. Así, para cada estado mental existe una descripción funcional tal que si un sistema la satisface se halla en ese estado mental en particular independientemente de que el sistema esté hecho de hidrocarburos o metales y silicatos.

Hay que señalar que a pesar de que hemos presentado, como suele hacerse, al funcionalismo como una forma de fisicalismo, la tesis nuclear del mismo no implica necesariamente el rechazo del dualismo: bien podría darse el caso de que un estado jugara determinado papel causal en la economía de un organismo o sistema sin que fuera él mismo físico o material. Para huir del dualismo es necesario sumar a esta tesis nuclear una adicional según la cual sólo se dan relaciones causales entre entidades materiales, en el contexto de los procesos físicos, una tesis adicional que puede entenderse ya como ínsita en la comentada relación de superveniencia. El funcionalismo, a diferencia del conductismo lógico y las teorías de la identidad, no es en sí mismo un planteamiento reduccionista, pues, por una parte, frente a éstas, el funcionalismo concede un importante grado de autonomía a la psicología frente a la neurofisiología al conceptualizar los estados y propiedades mentales como estados y propiedades funcionales no directamente físicas y caracterizables así en un vocabulario cuyo nivel de abstracción se hallaría por encima del propio del discurso sobre lo físico, respecto del cual tendría una autonomía que vendría dada por el hecho de que lo que cuenta en la descripción de un estado mental son antes las relaciones causales o funcionales en las que el mismo se vea implicado que el substrato material que las sustente. Las propiedades y estados mentales son para el funcionalista propiedades y estados funcionales y no directamente propiedades y estados físicos, aunque supervengan a éstos. Por otra parte, frente al conductismo lógico, el funcionalismo acepta que la mente tiene una estrecha relación con la conducta, pero se muestra en este punto más flexible que aquél en tanto concibe los estados mentales como causas internas de la conducta o las disposiciones para la misma en lugar de concebirlos como idénticos a éstas.

Presentado breve y esquemáticamente el marco general del funcionalismo, pasemos a exponer de forma igualmente compendiosa los rasgos principales de las diferentes formas de funcionalismo: el computacional, el analítico, el teleológico y el homuncular.

La primera teoría funcionalista en filosofía de la mente se debe a Hilary Putnam, que desarrolló sus tesis en una serie de artículos publicados en la década de los sesenta, de entre los cuales, tal vez el más influyente haya sido “Minds and machines”, texto de

1960 que puede ser leído hoy como el documento fundacional del funcionalismo computacional. Brevemente, esta prístina tendencia funcionalista se caracteriza por el trazado de la recurrente analogía entre mente y programa de ordenador. Como Manuel García-Carpintero ha señalado (García-Carpintero, 1995: 43-76), dicha analogía tiene a su base una conceptualización de la noción de programa de ordenador como descripción funcional, como una descripción de un proceso causal que incluiría referencias a estados internos. Así, en la propuesta del Putnam de los sesenta, los estados mentales aparecen como estados computacionales especificados en tales descripciones funcionales, y más concretamente como estados especificables mediante descripciones funcionales análogas a las necesarias para describir el estado de una máquina de Turing (de aquí que resulte habitual encontrar que se denomina a esta perspectiva teórica *funcionalismo de máquina*). En el funcionalismo computacional, en muy resumidas cuentas, los estados mentales se identifican con estados computacionales o funcionales del sistema en el que éstos se implementan, y la relación entre los estados mentales y los físicos (es decir, el meollo del tradicional problema mente-cuerpo) es planteada en paralelo a la habida entre los estados lógicos y los estructurales de una máquina de Turing.

El funcionalismo analítico, desarrollado por autores como Armstrong y Lewis, ofrece un análisis funcional distinto de los conceptos mentales. A diferencia del funcionalismo computacional, el analítico no se postula como una hipótesis científica, sino como un análisis de nuestro concepto de lo mental, como un análisis de los conceptos de estados mentales según el cual éstos "equivalen esencialmente al concepto de aquello que presenta un tipo particular de papel causal" (Priest, 1991: 179 de la traducción). Lewis mantiene que las teorías científicas constituyen descripciones funcionales dentro de las cuales los conceptos teóricos quedan definidos por los papeles causales que se les asigna dentro de la teoría que sea el caso. Pues bien, los términos de la psicología popular o de sentido común,[14] términos tales como deseo, esperanza, recuerdo, creencia, intención, etc., son concebidos por Lewis como términos teóricos definidos funcionalmente por el papel causal que desempeñan dentro de dicha teoría.

Por su parte, el funcionalismo teleológico, influido por la biología evolucionista, inserta en el marco funcionalista nociones tales como las de fin o propósito. Desde este punto de vista, un determinado estado mental es tal en función del papel que cumple de cara a propiciar que el organismo en el cual se da alcance sus metas o propósitos. Según

[14] Para una concisa introducción a esta noción vid. Rudder Baker (1999).

la conceptualización de la noción de función que ofrece esta opción teórica, los estados mentales cumplen determinadas funciones porque tienen la capacidad de causar determinados acontecimientos y porque existen dado que tienen esa capacidad. Así, esta forma de funcionalismo no atiende meramente a lo que los estados mentales hacen, sino asimismo a lo que se supone que deben hacer, es decir, a aquello que están evolutivamente diseñados para hacer. En definitiva, los estados mentales son desde el punto de vista del funcionalismo teleológico estados funcionales, pero un estado funcional es en este marco teórico un estado dotado de función en sentido biológico.

El funcionalismo homuncular, por su parte, es una propuesta teórica desarrollada por Daniel Dennett (Dennett, 1975) que otros autores –como William Lycan– adoptarían posteriormente –integrando en su caso perspectivas teleológicas–. Como Lycan explica (1987, cap. 4), Dennett se habría apoyado para la formulación de su funcionalismo homuncular en la metodología de determinados proyectos de inteligencia artificial en los cuales una labor caracterizada intencionalmente (¿cómo lograr que mi sistema comprenda enunciados?) se descompone en subtareas igualmente caracterizadas intencionalmente pero cada vez más simples (¿cómo lograr que mi sistema diferencie sujetos de predicados?), que a su vez se descomponen en problemas más simples, y así sucesivamente hasta alcanzar un nivel en el que las descripciones de las rutinas seguidas refieren obviamente a procesos mecánicos. Dennett ha sido ciertamente explícito al hablar del modo en que invita a contemplar el nivel intencional subpersonal en paralelo al personal: “The theory of content I espouse for the whole person I espouse all the way in” (Dennett, 1994a: 508).

*El eliminativismo*

De entre las perspectivas fisicalistas acerca de la ontología de la conciencia quizá la más dura o radical sea el *eliminativismo*. No obstante, el mismo se presenta en diferentes variedades, y al propio Dennett se ha atribuido en ocasiones una suerte débil de eliminativismo consistente en afirmar que la noción consuetudinaria de conciencia carece de un referente real o se halla contaminada e inflada por intuiciones que la convierten en algo más complejo e inasible que aquel fenómeno real que una ciencia de la conciencia debiera explicar. En las variantes más duras del eliminativismo, sin embargo, la propia existencia de la conciencia es negada: dicha noción no podría ser sustituida por otra más operacionalizable, sino sólo eliminada con el resto del vocabulario mentalista,

en el lugar del cual se erguiría el aparato descriptivo y explicativo de una futurible neurociencia. El eliminativismo fue desarrollado en la década de los ochenta por el matrimonio Churchland, pero sus antecedentes pueden ser rastreados en la filosofía norteamericana de la década de los cincuenta. Así, por ejemplo, Wilfrid Sellars (Sellars, 1956) propondría una de las ideas que posteriormente ubicaran los Churchland en el núcleo de su planteamiento eliminativista, la idea de que nuestras nociones acerca de nuestra vida mental pueden derivar, antes que de nuestro acceso al mundo privado de las entidades mentales a que referirían los términos de la psicología popular, de nuestra herencia cultural. Se trata de una idea central para el eliminativismo dado que, partiendo de ella, nuestro discurso acerca de entidades mentales puede ser visto no como designando entidades a las que tenemos de hecho un acceso privado y privilegiado, sino como basado en un marco teórico heredado culturalmente. Otro influyente filósofo americano cuyas aportaciones prefiguraran las del eliminativismo fue Willard van Orman Quine. En la que muchos consideran su obra principal (Quine, 1960), Quine propondría que la referencia de las palabras que usamos para denotar estados mentales podría resultar mejor tratada al ser vinculada con estados fisiológicos que con supuestas entidades mentales. De estas dos ideas parte el eliminativismo de los Churchland, que puede resumirse de forma muy sencilla: los términos que constituyen la psicología popular son términos teóricos pertenecientes a una teoría falsa y acabarán siendo eliminados por el avance de las neurociencias. Es decir, desde el punto de vista eliminativista, cuando digo "recuerdo emocionado el día en que te conocí", mi aserto sería irreferencial mientras no sean en él adecuadamente sustituidos los términos de la psicología popular por términos neurofisiológicos apropiados. No se trataría, en definitiva, de reducir el lenguaje mentalista al de las neurociencias, sino de eliminarlo, dado que es irreferencial y forma parte de una teoría falsa. El lenguaje mentalista no es, desde este punto de vista, un lenguaje teórico deficiente, sino vacío: no designa nada, es decir, nunca han existido entidades tales como los deseos, las creencias, las sensaciones o las imágenes mentales. El modo en que este planteamiento se aplica al debate en torno al problema de la conciencia resulta obvio: todos los términos mentalistas utilizados en dicho debate son *flatus vocis*. Así, no existe, tan siquiera, algo como eso que pretendía designarse con la palabra «conciencia», de forma que ésta debiera ser eliminada junto con el resto de los términos que incluye la psicología popular. No obstante, a esta forma radical de eliminativismo precedería una discusión en torno a la diferencia entre la noción de reducción y la de eliminación. Ambas nociones se encontraban entreveradas en el artículo

pionero de Richard Rorty "Mind-body identity, privacy, and categories", de 1965. En este artículo se palpaba una tensión entre la idea de que las sensaciones conscientes sencillamente no existen y la idea de que efectivamente existen pero no son más que procesos neurofisiológicos. Un artículo posterior de William Lycan y George Pappas, intitulado "What is eliminative materialism?", pondría definitivamente las cartas sobre la mesa: o se sostiene que las nociones mentalistas de la psicología popular sencillamente son irreferenciales y no designan por tanto nada real, permaneciendo en el bando propiamente eliminativista, o se sostiene, desde la perspectiva tradicional de un materialismo reduccionista como el propio de las teorías de la identidad, que las nociones mentalistas efectivamente designan algo, pero algo diferente de aquello que dentro del marco de la psicología popular vino considerándose que designaban, de forma que deben ser, para corregir su referencia, reducidas al vocabulario de las neurociencias. La primera de las opciones sería poco después denominada ontológicamente radical, la segunda ontológicamente conservadora (vid. Savitt, 1974). Partiendo de esta segunda, nociones propias de un léxico teórico concreto podrían traducirse, acaso con alguna que otra modificación, reubicándose en el marco de otro vocabulario teórico, mientras que partiendo de la primera tal procedimiento sería completamente inviable. En cualquier caso, la diferencia entre ambas opciones estribaría en que, en el marco de un fisicalismo propiamente eliminativo, nuestros conceptos cotidianos referidos a estados o eventos mentales de ningún modo pueden reducirse o identificarse con procesos o eventos neurobiológicos, dado que forman parte de una concepción teórica completamente errónea. Esta línea radical sería desarrollada en los ochenta por los Churchland, como señalábamos, y, asimismo, por Stephen Stich y Georges Rey.

Al igual que en el caso de las teorías de la identidad y el funcionalismo, el eliminativismo no es una planteamiento filosófico destinado exclusivamente a ofrecer una caracterización adecuada de la naturaleza de la experiencia consciente, sino que sus propuestas conforman, cabría decir, un paradigma en filosofía de la mente que cubre el espectro completo de temas debatidos en el área. No obstante, Georges Rey y Kathleen Wilkes, entre otros, han expuesto con claridad (Rey, 1983; 1988; Wilkes, 1984; 1988; 1995) tanto el móvil como las consecuencias de un tratamiento elimiativista del núcleo del problema de la conciencia, esto es, la conciencia fenoménica. En sus propuestas, nuestra capacidad para imaginar el modo en que los procesos neurobiológicos o computacionales propuestos por algunas teorías explicativas de la experiencia consciente podrían de hecho tener lugar sin ocasionar experiencia conciente (el denominado *explana-*

*tory gap*) o, complementariamente, la incapacidad de esas teorías para capturar nuestra noción cotidiana de experiencia consciente, podría perfectamente deberse a que eso que llamamos experiencia consciente no sea sino un mero remanente espurio de un puñado de desorientadoras aunque persistentes intuiciones cartesianas. Este eliminativismo respecto de la conciencia fenoménica no ha dejado de resultar una postura un tanto anti-intuitiva, motivo por el cual no resulta extraño encontrar que en determinados lugares se viertan sobre ella las más acerbas de las críticas, llegando a ser descrita como "algo que implica la desvergüenza de negar lo evidente" (Rodríguez González, 2010: 184).

> *Hasta aquí hemos hablado de teorías cuyo propósito es el de definir qué es la conciencia y cuál ha de ser el lugar que ocupe en nuestra concepción de la realidad. En el próximo capítulo nos aproximaremos a las que han tratado de disponer lo necesario no ya para esbozar una definición de «conciencia», sino una explicación de los motivos y mecanismos de la existencia de la experiencia consciente. El resto de la tesis puede leerse como un intento de a) cotejar la estrategia definitoria y la explicativa y b) despejar el camino para el desarrollo de esta última.*

# CAPÍTULO 5

## CÓMO ENCAJA LA CONCIENCIA. MAPA DE LAS PROPUESTAS EXPLICATIVAS

Son varias las posibilidades con las que contamos de cara a trazar un esquema de los diferentes intentos explicativos hasta la fecha ensayados por científicos y filosóficos con vistas a resolver el problema de la conciencia, intentos destinados a ofrecer respuesta no a la pregunta acerca de qué es la conciencia y qué lugar ha de ocupar, por tanto, en nuestra concepción de la realidad, como en el caso de las perspectivas ontológicas, sino a la pregunta acerca de cómo y por qué surge o existe la conciencia. Hemos optado aquí por clasificar estos intentos en tres bloques: teorías cognitivas, teorías representacionales y teorías neurobiológicas.[1] Con este esquema introductorio nos proponemos, por una

[1] No entraremos a comentar aproximaciones explicativas más "exóticas", como las ensayadas desde la teoría del caos y los fractales (Gu, Meng, Shen & Cai, 2003; King, 2003; MacCormac & Stamenov, 1996; Van Gelder, 1999a; Walling & Hicks, 2003), los sistemas dinámicos no-lineales y sistemas disipativos (King & Pribram, 1995; Pitkänen, 2003; Roederer, 2003; Scott, 1995; Vitiello, 2003), la teoría de supercuerdas (Godfroid, 2003) o la mecánica cuántica. Dentro de esta clase de aproximaciones "exóticas", la que viene disfrutando de mayor difusión es esta última. Sin embargo, y a pesar del reciente entusiasmo de Hameroff & Penrose (2014) a causa del descubrimiento de vibraciones cuánticas en microtúbulos por el grupo de Anirban Bandyopadhyay –observaciones que, supuestamente, vendrían a confirmar la teoría Orch-OR, elaborada a mediados de los noventa por Penrose y defendida y desarrollada desde entonces por ambos como un modelo que sitúa el origen de la conciencia en computaciones cuánticas acaecidas en los microtúbulos, en un proceso de autoorganización en los intersticios de la geometría espacio-temporal, a escala de Planck, vinculado, por tanto, con la estructura fundamental del universo (sic)–, la opinión más extendida en los *Consciousness Studies* sigue siendo hoy, como lo era hace dos décadas, que las teorías cuánticas de la conciencia –presentadas por sus valedores como "the most comprehensive, rigorous, and successful theor[ies] of consciousness ever put forth" (Hameroff, 2014: 148)– tratan de resolver algo que parece un misterio echando mano de algo que lo parece aun más: "some pundits have suggested that the spate of publications over the last decade on the possibility that quantum mechanical principles are needed to explain consciousness arise from the conviction that two such fundamental mysteries must be related!"(Bennett, 1997: 144). [Antonio Damasio, en su último libro hasta la fecha, expresa la misma idea en los mismos términos, pero no cita a Bennett (Damasio, 2010: 14 del original; 35 de la traducción)]. No obstante, huelga subrayar que, a pesar de lo peregrina que –en vista del *corpus* de evidencias que desde la

parte, clarificar el sentido en que cabe distinguir teorías ontológicas y explicativas –a pesar de que, como sugeríamos, todo intento explicativo se halla ineluctablemente vinculado a determinados supuestos ontológicos– y, por otra, preparar el terreno para la crítica de la idoneidad de los términos en que ha venido planteándose el debate en torno al problema de la conciencia que emprenderemos en la segunda parte. Situaremos el punto de partida de dicha crítica en la idea de que el problema que las teorías explicativas abordan es el verdadero problema de la conciencia, motivo por el cual, de no acercarnos previamente a las mismas, quedaría nuestra crítica renca y como flotando en el aire.

### 1._Teorías cognitivas

Las teorías de la conciencia de las que vamos a tratar en este apartado intentan delimitar el lugar que la conciencia ocupa dentro del entramado de capacidades mentales que constituyen el sistema cognitivo humano. Desde el punto de vista de los teóricos cognitivos, una vez delimitado ese lugar, la conciencia habría sido explicada y su existencia justificada, dado que ello implicaría especificar la función que desempeña. Estas teorías pueden ubicarse dentro de la tradición de la psicología cognitiva, siendo así que propuestas teóricas pioneras en dicha tradición, como el ya mencionado modelo del filtro de la atención selectiva de Broadbent, contribuirían a sentar las bases del marco teórico dentro del cual han venido desarrollándose las teorías cognitivas de la conciencia. El modelo de Broadbent postulaba, como vimos, la existencia de un canal de capacidad limitada dentro del cual podía procesarse de forma consciente la información que ingresaba en el sistema cognitivo. Broadbent pretendía, sencillamente, construir un modelo teórico capaz de dar cuenta de sus resultados experimentales en tareas de escucha dicótica. No obstante, su idea de un canal de capacidad limitada derivó en la década de los sesenta en la de un procesador central de capacidad limitada, concebido como ejecutivo central o sistema supervisor. A este subsistema, emplazado en lo más alto del edificio cognitivo, se le atribuía la ejecución y control de algunas de las capacidades cogniti-

neuropsicología y la neurofisiología viene apuntando en dirección contraria– pueda resultar la invitación a abandonar la doctrina neuronal (Gold & Stoljar, 1999) y descender al nivel cuántico en busca de la conciencia, descartar de plano la posibilidad de que a determinados sectores de la física no clásica les quepa contribuir en algún sentido a la explicación de la experiencia consciente y sobrestimar las dotes adivinatorias propias no parecen constituir, desde según qué perspectiva, instancias actitudinales diferencialmente categorizables. Con todo, algunos tienen muy claras las perspectivas de futuro de la teoría Orch-OR: "Pixie dust in the synapses is about as explanatorily powerful as quantum coherence in the microtubules" (Churchland, 1998: 121).

vas de alto nivel, como la toma de decisiones, la elaboración de planes o la supervisión y coordinación de subsistemas situados en un nivel inferior de la jerarquía de procesamiento.[2] A pesar de que durante la década de los sesenta la noción de conciencia no se asociara explícitamente a este ejecutivo central, en la subsiguiente destacados psicólogos no tuvieron reparos a la hora de definir a la conciencia como una forma de procesamiento de información llevada a cabo por un subsistema cognitivo central de capacidad limitada (Posner & Warren, 1972). Al igual que en muchas teorías cognitivas contemporáneas, estos sistemas de procesamiento central de las décadas de los sesenta y los setenta fueron presentados como estructuras dinámicas vertebradoras de la actividad del sistema cognitivo, unas estructuras cuyo cometido estribaba en integrar la información y las funciones de otros subsistemas y garantizar así un procesamiento flexible y controlado como el requerido para el desarrollo de la conducta adaptada propia de los sujetos no patológicos.[3] Una conducta tal requiere en muchas ocasiones superar los automatismos propios de procesos cognitivos y conductuales no controlados conscientemente. Así, podemos, por ejemplo, conducir de forma automática hasta que surge una situación problemática para la que no existe una respuesta típica previamente aprendida y automatizada y necesitamos actuar prestando atención a nuestros movimientos y tratando de controlarlos adecuadamente con arreglo a determinadas metas. Este tipo de situaciones fueron las que captaron la atención de los herederos de la noción cognitivista de ejecutivo central en la década de los ochenta. El modelo de Norman y Shallice fue el primero en afrontar la tarea de elucidar el modo en que el sistema cognitivo puede enfrentarse a la solución de problemas que requieren de una atención consciente dirigida voluntariamente. Dicho modelo (Norman & Shallice, 1980) parte de una idea similar a la de *script* (Schank & Abelson, 1977), pues su supuesto fundamental es el de que muchas conductas responden a lo que los autores denominan *esquemas de acción*, estructuras que controlan procesos atencionales automáticos y que pueden dispararse como consecuencia de un insumo sensorial o del output de otro esquema. La noción de ejecutivo central aparecería aquí bajo el nombre de *Supervisory Attentional System* (SAS), un mecanismo cognitivo que actuaba como un sistema de control de alto nivel puesto en marcha ante

[2] En este sentido, los planteamientos cognitivos retomarían la noción de conciencia de Claparède (1933), que se refería a ella como un "aparato de control".

[3] El modelo de la memoria de trabajo que Alan Baddeley perfilara en las décadas subsiguientes postuló ya en ésta de los setenta la existencia de uno de los sistemas de ejecutivo central de mayor pujanza en psicología cognitiva (Baddeley & Hitch, 1974). En dicho modelo, el contenido de la conciencia coincidía con el de la memoria de trabajo, un planteamiento que ha ejercido una enorme influencia en las teorías cognitivas de la conciencia que comenzaran a elaborarse a mediados de la década de los ochenta.

situaciones no rutinarias en las cuales los esquemas de acción almacenados y automatizados pueden resultar inadecuados. El SAS contenía sistemas de planificación de carácter general que operaban de una forma más flexible pero menos rápida que los automáticos esquemas de acción, los cuales podían, de hecho, resultar inhibidos por acción del SAS en caso de que fuera inapropiado que uno de ellos se ejecutara en un determinado contexto.

Estos pioneros modelos de ejecutivo central ofrecieron una descripción global de la arquitectura del sistema cognitivo humano en la cual aparecían diferenciados subsistemas encargados de llevar a cabo diferentes tareas: los órganos de los sentidos constituían la puerta de entrada de la información al sistema cognitivo y sus insumos eran procesados en paralelo a través de diferentes vías o canales, la atención realizaba la selección o filtrado de la información procesada en esos canales, la memoria de trabajo la mantenía activada y accesible y un sistema central modulaba y supervisaba globalmente el procesamiento de información llevado a cabo por el sistema cognitivo.

En esta época, el creciente interés en la adecuada demarcación de segmentos diferenciados del sistema cognitivo destinados a desempeñar diferentes funciones encontraría soporte teórico en la influyente obra de Jerry A. Fodor *The Modularity of Mind*, que acertó a ofrecer, en el momento adecuado, una robusta argumentación en favor de la necesidad de postular "many fundamentally different kinds of psychological mechanisms" (Fodor, 1983: 1 del original, 19 de la traducción). Nos hallamos, pues, en una época en la que comienzan a trazarse mapas de una arquitectura cognitiva para la cual muchos buscaban un componente central dotado de un estilo de procesamiento diverso del que se atribuía al resto de subsistemas del aparto cognitivo. Se plantea de este modo la existencia de diferencias estructurales, encarnadas en jerarquías de procesamiento, y diferencias en el propio estilo de procesamiento, caracterizadas en términos de la famosa dicotomía automático/controlado (Posner & Snyder, 1975): mientras el procesador central implicaría control voluntario y actuaría de forma lenta pero flexible, los procesadores periféricos operarían de forma rápida y automática, pero rígida. En gran cantidad de modelos, esta arquitectura y, particularmente, estos estilos de procesamiento se asimilaron –en ocasiones de forma implícita–, respectivamente, a un procesamiento consciente e inconsciente de información. Esta concepción estructural de un sistema cognitivo que albergaría subsistemas con diversos estilos de procesamiento (rápido/lento, rígido/flexible, automático/controlado) se halla a la base de las teorías cognitivas de la conciencia contemporáneas. No obstante, fueron pocos los que en aquella épo-

ca se decidieron a ofrecer descripciones funcionales de ese conjetural subsistema central al que suponían un estilo de procesamiento lento, flexible y controlado (consciente),[4] lo cual no ha de extrañar cuando el mismo fue a menudo descrito de forma un tanto inespecífica y concebido como una parte del sistema cognitivo que, a pesar de que resultara intratable y misteriosa, era necesario postular a la vista de la conducta exhibida por sujetos normales tanto en situaciones cotidianas como en diferentes contextos experimentales. En este sentido, una figura tan influyente como la de Fodor subrayó en esta época la dificultad que sin duda habría de implicar la tarea de ofrecer una caracterización apropiada de este componente central del sistema cognitivo.

Pueden considerarse los planteamientos de Philip N. Jonhson-Laird (1983a; 1983b) como antecedente y referente fundamental de las teorías cognitivas de la segunda mitad de los ochenta. Jonhson-Laird, en la línea de la tradición inaugurada por Broadbent, comparó a la conciencia con el sistema operativo de un ordenador, que distribuye recursos computacionales a las diferentes tareas en curso del mismo modo que aquélla distribuye recursos atencionales. La conciencia, en términos de arquitectura cognitiva, ocupó el lugar central y más alto del edificio cognitivo en todos los modelos que, hasta el de Jonhson-Laird, inclusive, trataron de ella de un modo u otro. No obstante, y curiosamente, la primera teoría cognitiva elaborada de forma sistemática y con la intención explícita de explicar cómo funciona y cómo se origina la conciencia, no situó el procesamiento consciente de información en lo más alto de la jerarquía cognitiva, sino en un nivel intermedio. Esta teoría apareció publicada en 1987 en *Consciousness and the Computational Mind*, de Ray Jackendoff. El núcleo de la misma consistió en asociar el procesamiento consciente de información con representaciones de nivel intermedio. Así, por ejemplo, en el caso de la percepción del lenguaje, cabría hablar de un primer nivel acústico de representación, que sería en el que operarían los sistemas cognitivos periféricos más próximos a los insumos recibidos a través de los oídos y analizados con arreglo a las propiedades físicas del estímulo, y un nivel abstracto de representación que operaría con categorías léxicas y sintácticas abstractas. Jackendoff vinculó el procesamiento consciente de información con un nivel de representación intermedio ubicado entre los dos señalados (un nivel que, en el caso del ejemplo propuesto, tendría que ver con representaciones de tipo fonológico) y desarrolló su modelo con una intención más

[4] Una notable excepción puede encontrarse en el modelo *Adaptive Control of Though*, de John Anderson (vid., v. g., Anderson, 1983).

humilde que la de construir una teoría completa acerca de *cómo* surge la conciencia: la intención de señalar *dónde*, en qué punto de la jerarquía del procesamiento de la información en nuestro sistema cognitivo emerge la experiencia consciente. Su hipótesis de que ese lugar se encuentra en un nivel intermedio de la jerarquía de procesamiento puede entenderse hoy como apoyada por una considerable cantidad de evidencias empíricas. Tomemos como ejemplo el sistema visual. Las lesiones al nivel más básico, en las áreas visuales primarias (V1), impiden que la información se haga consciente, dado que bloquean su curso hacia el nivel intermedio hipotetizado en este marco teórico. Algo análogo sucede con las lesiones en el propio nivel intermedio (áreas extraestriadas), que producen ceguera para atributos específicos, como la acromatopsia (lesiones en V4) o la acinetopsia (lesiones en V5). Sin embargo, las lesiones en niveles superiores del procesamiento visual no suprimen la experiencia consciente, sino el reconocimiento de objetos (producen agnosias); es decir, están asociadas a la incapacidad para trazar vínculos entre perceptos y conceptos, una incapacidad a pesar de la cual el paciente puede experimentar conscientemente y describir los atributos perceptivos de los objetos que le son presentados.[5]

Un año después de que Jackendoff publicara *Consciousness and the Computational Mind*, Bernard J. Baars presentó en *A Cognitive Theory of Consciousness* la primera versión de la teoría cognitiva de la conciencia más exitosa e influyente hasta la fecha: la *teoría del espacio de trabajo global*. La teoría de Baars parte de la habitual asunción según la cual en nuestro sistema nervioso central hay constantemente una enorme cantidad de procesos de manipulación de información en curso que son relevantes para nuestra conducta adaptada, y a pesar de ello, nosotros sólo llegamos a ser conscientes de una ínfima parte de esa información. Si la inmensa mayoría de esa información puede ser perfectamente manejada de forma inconsciente, ¿qué utilidad tiene que una pequeña parte de la misma se haga consciente? La respuesta de Baars consiste en definir la conciencia como la puerta hacia las fuentes de control y conocimiento inconscientes del sistema nervioso, una puerta que, al abrirse, permite la integración de gran cantidad de procesos y funciones realizadas en procesadores neuronales altamente especializados que de otro modo tendrían lugar de forma aislada e independiente. La conciencia es así, en la propuesta de Baars, un medio para el acceso, diseminación e intercambio de in-

[5] En nuestros días, Jesse Prinz ha defendido una teoría de nivel intermedio que supone un refinamiento de la de Jackendoff (vid. Prinz, 2000; 2001; 2007).

formación, y para la puesta en práctica de funciones de coordinación y control global. En un sistema de procesadores neuronales altamente especializados que realizan sus tareas en paralelo y de forma relativamente autónoma e independiente, la conciencia vendría a posibilitar la señalada coordinación y control global propiciando un intercambio centralizado de información en virtud del cual procesadores especializados y de otro modo silentes logran hacer globalmente accesible los productos de su actividad.

> Coordination and control may take place by way of a central information exchange, allowing some processors (...) to distribute information to the system as a whole. (Baars, 2005: 46).

Otra de las teorías cognitivas de la conciencia más famosas y discutidas es el *Modelo de Versiones Múltiples* (MDM, por sus siglas en inglés), de Daniel C. Dennett. Dedicaremos a este modelo una sección en la tercera parte de esta tesis y ofreceremos por tanto en este punto sólo una breve introducción al mismo. A pesar de que Dennett (1978) había tratado ya de avanzar hacia una teoría cognitiva de la conciencia, no sería hasta 1991 que elaborara el MDM (Dennett, 1991a: cap. 5), que recientemente ha preferido describir haciendo uso de la metáfora de la "fama en el cerebro" (Dennett, 2005: caps. 6 y 7). Dennett opone su modelo a la idea –según él endémica– de un Teatro Cartesiano de la conciencia, el lugar en que se proyecta la película de la conciencia, un espacio central al que todo llega junto y a la vez para una clara y distinta presentación unitaria ante la audiencia, ante el homúnculo cartesiano. Dennett nos invita a deshacernos de la idea de este centro advirtiendo que en el cerebro la información es procesada masivamente en paralelo a través de diversas vías en las que la misma es sucesivamente elaborada e interpretada, no habiendo en el sistema nervioso central ningún punto al que lleguen acendrados los resultados de dicho procesamiento. El procesamiento serial que nuestra ordenada experiencia consciente sugiere es confrontado en el marco del MDM a una vorágine paralela de procesamiento de información llevada a cabo por una enorme cantidad de módulos en un proceso en el que la información es reelaborada a la luz de nuevas contingencias, es decir, reelaborada y revisada con vistas a su integración coherente con la nueva información que constantemente accede al sistema nervioso central. Nos hallaríamos ante un flujo masivo y paralelo de procesos simultáneos de fijación y transformación de contenido. De esta manera, dispondríamos en todo momento de gran variedad de versiones de la misma información, y la ilusión de que dicha información ingresa serialmente en el Teatro Cartesiano se debería a la ocurrencia de recurrentes

sondeos que favorecen unas versiones de dicha información en detrimento de otras. No puede fijarse ni el lugar ni el momento en el que algo se hace consciente, y el resultado de esos procesos paralelos de elaboración e interpretación de la información, el resultado que llamamos conciencia, es la generación de un flujo narrativo estrechamente vinculado con nuestra capacidad lingüística y mnemónica. Nuestra experiencia consciente sería pues el resultado de multitud de procesos interpretativos operados sobre una gran cantidad de versiones de los mismos contenidos, las cuales se disputarían el predominio, alcanzándolo aquéllas que se muestran capaces de monopolizar recursos y repercutir en la conducta y la memoria.

La última teoría cognitiva de la conciencia que comentaremos se encuadra, a diferencia de sus precursoras simbolistas de las décadas setenta y ochenta, en el paradigma conexionista, en la modelización computacional de funciones mentales mediante redes neuronales artificiales. También las teorías de Baars y Dennett hacen uso del utillaje teórico conexionista, pero mientras en ellas se prolonga la teorización arquitectónica que predominara en las décadas previas, en la que a continuación presentamos ese énfasis arquitectónico se desplaza hacia las propiedades de las representaciones implementadas en redes neuronales artificiales, pues de tales propiedades depende, según esta teoría, que una representación pueda o no hacerse consciente. A diferencia de lo que sucedía en el paradigma simbólico, la actividad representacional de la mente no consiste en el paradigma conexionista en el procesamiento serial de símbolos discretos, estructurados sintácticamente y manipulados de acuerdo con reglas ajenas al contenido representado, sino que la noción de representación se reformula en el paradigma conexionista en términos de patrones de activación y conexión entre las unidades de redes neuronales artificiales y la de computación se redefine en términos de transiciones entre tales patrones. Dentro de este marco teórico, Axel Cleeremans ha propuesto la *tesis de la plasticidad radical* (Cleeremans, 2008). Según la misma, las representaciones de más calidad en la dinámica de una red neuronal, esto es, aquéllas cuyos patrones de activación son más diferenciados, fuertes y estables (Cleeremans, 2005: 84), tendrán más posibilidades de participar en la vida consciente del sistema. El aprendizaje es, pues, una de las claves de la teoría de Cleeremans. Con él, en el entrenamiento por exposición del sistema a su entorno, las representaciones implementadas en las redes neuronales mejoran su calidad. Cleeremans defiende que también un proceso meta-representativo puede aumentar la calidad de una representación, ofreciéndole mayores posibilidades de interve-

nir en la vida consciente del sistema. Estas meta-representaciones aumentarían la calidad de las representaciones de primer orden –derivadas del aprendizaje del sistema por exposición a su entorno– al perfilarlas en el proceso de poner en relación sus características con las de otras representaciones de primer orden. Curiosamente, esta teoría cognitiva llaga a una conclusión opuesta a las de sus precursoras, que identificaban, como vimos, procesamiento automático con procesamiento inconsciente. Según Cleermans, como indicábamos, una representación tiene mayores posibilidades de concurrir en la conciencia de un sistema cuanto mayor sea su calidad. Dicha calidad aumenta con el aprendizaje, y precisamente el aprendizaje tiende a propiciar la automatización de determinadas representaciones. Según los estándares de Cleeremans, éstas serían representaciones de alta calidad y podrían intervenir con facilidad en la vida consciente del sistema.

## 2._Teorías representacionales

Las teorías representacionales de la conciencia son más recientes que las primeras teorías cognitivas y su objeto es más específico: el del modo en que el aspecto representacional o intencional de lo mental podría explicar la existencia de su aspecto fenoménico. Al igual que en el caso de las teorías cognitivas, las teorías representacionales se hallan insertas en un marco teórico funcionalista tácita o explícitamente próximo al computacionalismo y a la psicología del procesamiento de información. Sin embargo, a diferencia de las teorías cognitivas, que fueron planteadas en términos en principio operacionalizables y susceptibles de enfrentarse a confutación empírica, las teorías representacionales intentan ofrecer la referida explicación en un plano puramente teórico, en abstracto, por así decir. Así, las teorías representacionales de la conciencia han venido siendo el terreno habitual para la discusión filosófica acerca de las posibilidades de explicar científicamente la conciencia y han sido generalmente formuladas por filósofos, mientras las teorías cognitivas lo fueron por psicólogos. Señalemos antes de pasar a trazar un mapa de las teorías representacioales que mientras a menudo se asume que las teorías cognitivas tratan –acaso subrepticiamente– a la conciencia fenoménica en términos de conciencia de acceso, las teorías representacionales, a pesar de haber dado pocos o ningún paso hacia una concreta operacionalización de sus nociones, abordan el problema de la conciencia *tout court*, esto es, se refieren propiamente a las posibilidades de ofrecer una explicación científica de la conciencia fenoménica –aunque bien es cierto

que, dada su falta de concreción y dados, principalmente, los escasos esfuerzos de operacionalización realizados por los teóricos representacionalistas, estas teorías podrían incluirse en una categoría a medio camino entre lo ontológico y lo explicativo.

Una enorme cantidad de teorías representacionales ha visto la luz en los últimos años. Esta proliferación responde al general consenso acerca de las mayores posibilidades de operacionalizar y explicar científicamente el aspecto representacional de los estados mentales. De este modo, dentro de diferentes paradigmas en las ciencias encargadas del estudio de lo mental, hemos ido viendo cómo una aproximación científica a fenómenos mentales tales como el aprendizaje (punta de lanza del paradigma conductista), la atención, la memoria (núcleos iniciales del desarrollo del computacionalismo en psicología) o la percepción (ariete del paradigma conexionista) parece poder ensayarse y acendrarse progresivamente sin topar con óbices insalvables, e incluso cómo máquinas adecuadamente programadas exhiben una conducta análoga a la exhibida por sujetos humanos ante tareas de razonamiento, reconocimiento, categorización, etc. No obstante, el modo de aplicar a semejante clase de máquinas la distinción entre un procesamiento fenoménicamente consciente de la información y otro inconsciente se presenta como un problema inabordable, y así, el modo de operacionalizar adecuadamente el aspecto fenoménico de los estados mentales ha permanecido envuelto en las más densas de las brumas. En este contexto, las teorías representacionales vendrían a defender que el aspecto representacional de lo mental, supuestamente operacionalizado con éxito por diferentes paradigmas en ciencias cognitivas, agotaría el ámbito de lo mental, con lo cual la tarea consistiría en confeccionar especulativos modelos capaces de hacer palmario el modo en que lo representacional constituye la base sobre la que se alza lo fenoménico y, de hecho, el material del que lo fenoménico se halla constituido. En otras palabras, dado que las posibilidades de ofrecer una explicación naturalista –esto es, factible dentro de un marco teórico que no resulte problemático desde el punto de vista de los aceptados en ciencias naturales– del aspecto representacional de la mente han venido mostrándose prometedoras mientras que no ha venido sucediendo lo mismo con las de pergeñar conato análogo con el aspecto fenoménico, las teorías representacionales han venido tratando de reducir éste a aquél.[6] Las teorías representacionales, en resumidas cuentas,

[6] Hay que matizar que, tal y como Crane (2003) y Chalmers (2004) han puesto de relieve, el representacionalismo no ha de ser necesariamente reductivo. Esto es, caben, por ejemplo, posturas que presenten como representacional cualquier forma de conciencia fenoménica pero añadan que el contenido intencio-

tratan de diluir la tradicional heterogeneidad entre el aspecto intencional de la mente, que incluiría los fenómenos mentales caracterizables en términos de su ser-acerca-de, y su aspecto fenoménico, que incluiría los denominados *qualia*, esto es, los fenómenos mentales o aspectos de los mismos caracterizables en términos del modo en que son sentidos, experimentados. Diluir esta heterogeneidad implica en este contexto elaborar un marco teórico en el que todos y cada uno de los aspectos de todos y cada uno de los estados mentales, y en particular este aspecto fenoménico tradicionalmente concebido como ajeno a la mecánica representacional de la mente, resulten explicables en términos de la capacidad de lo mental para representar o ser-acerca-de. El objetivo de las teorías representacionales es, pues, el de extender el tratamiento de los estados intencionales al de los fenoménicos demostrando que ambos pueden ser comprendidos en términos de procesos representacionales, con lo cual, todas las teorías representacionales de la conciencia plantean que la conciencia fenoménica consiste, de algún modo, en una cierta clase de contenido representacional, el cual ocuparía un lugar concreto en la arquitectura funcional de la mente.[7]

*Teorías representacionales de primer orden*

Un primer grupo de explicaciones representacionales de la conciencia lo integrarían aquellos planteamientos que presentan el aspecto fenoménico de lo mental en términos de representaciones de primer orden. Fred Dretske y Michael Tye son los más destacados teóricos de este tipo de aproximación, aunque sus antecedentes pueden ras-

---

nal no puede ser caracterizado sino en referencia a su carácter fenoménico y que por tanto reducir la conciencia fenoménica a la economía representacional de la mente resultaría circular. No obstante, el representacionalismo, en el sentido relevante del término, está motivado por una asunción según la cual la intencionalidad resulta menos problemática que la fenomenalidad desde un punto de vista materialista y, por tanto, la opción más razonable para el materialista –ha venido presumiéndose– sería tratar de mostrar cómo lo fenoménico se reduce a lo intencional.

[7] No necesitamos hacer uso aquí de la habitual distinción entre representacionalismo puro o radical (según el cual toda representación, por el mero hecho de serlo, implica conciencia fenoménica), fuerte (según el cual las representaciones necesarias para alcanzar la conciencia fenoménica han de poseer una serie de propiedades distintivas) y débil (una postura que, meramente, consistiría en aceptar que los estados fenoménicamente conscientes poseen contenidos representacionales, cosa que, como resulta obvio, no implicaría necesariamente la defensa de un representacionalismo reductivo: cabe perfectamente que, a pesar de poseer contenidos representacionales, los estados fenoménicamente conscientes se caractericen por un extra no representacional dotado de un estatuto ontológico netamente heterogéneo respecto de lo intencional), dado que el primero es una teoría sin defensores –aunque no es infrecuente que se acuse a Thau (2002) de sostenerla– y el tercero ofrece antes descripciones fenomenológicas que explicaciones en el sentido que nos ocupa en este capítulo. Podemos pues ahorrarnos la etiqueta "fuerte" al hablar de representacionalismo, dado que el resto de representacionalismos o bien no son propiamente teorías explicativas o bien lo son pero nadie las defiende.

trearse en Anscombe (1965), Hintikka (1969), Kraut (1982), Lewis (1983a), Lycan (1987; 1996), Harman (1990) o Shoemaker (1994a). En sus teorías, las propiedades fenoménicas de los estados mentales aparecen descritas como contenidos representacionales, de tal modo que las diferencias entre las cualidades experimentadas al escuchar un do y un re, por ejemplo, serían explicadas en referencia a las diferencias habidas entre las propiedades objetivas representadas en cada caso. No hay en la experiencia consciente, desde este punto de vista, nada más allá del modo en que ella representa el mundo o el estado del cuerpo del sujeto. Los proponentes de este tipo de planteamiento sostienen, por otra parte, que un estado mental fenoménicamente consciente ha de poseer la capacidad de guiar la conducta mediante su influencia en las creencias y el razonamiento práctico. Así, la experiencia consciente es concebida por estos autores como el *output* de los diversos sistemas propioceptivos y sensoriales periféricos, un *output* a disposición de los sistemas cognitivos encargados de la formación de creencias, la generación de planes y el control del comportamiento. La conciencia fenoménica, definida como una forma de contenido intencional, queda así incardinada en la economía funcional del organismo. No obstante, dicho contenido no es entendido por los defensores de las teorías representacionales de primer orden como conceptual o proposicional: la conciencia fenoménica es, según ellos, una forma de representación capaz de intervenir en procesos de formación de creencias y control de la conducta, pero no una forma de representación que pueda ser capturada en conceptos y estructurada en proposiciones. La defensa de este carácter no conceptual de la clase de representaciones en que basan su propuesta adopta con frecuencia la siguiente forma: nuestro trato cotidiano e irreflexivo con los objetos que nos rodean y constituyen es un trato con las cualidades que los mismos poseen antes que con los conceptos mediante los cuales las describimos, de donde cabe colegir que nuestras experiencias conscientes se encontrarán antes referidas a dichas cualidades que a nuestras descripciones de las mismas (vid., v. g., Tye, 2000: 60-61).

Varias acotaciones serán sin duda de interés de cara a delimitar el carácter de esta clase de teoría representacional. Por una parte, como se desprende del ejemplo que propusiéramos al aludir a las diferencias entre las cualidades experimentadas al escuchar un do y un re, explicadas dentro de este marco teórico en virtud de las habidas entre las propiedades objetivas representadas en cada caso, estas teorías representacionales no conciben la representación como una mediación entre, verbigratia, lo percibido y la percepción, sino que se asume un contacto directo del organismo con el mundo y consigo

mismo. Así, por ejemplo, el contenido de una experiencia perceptiva no sería una entidad mental que haría las veces de intermediario entre dicha experiencia y el objeto o las propiedades percibidas, sino que éstas mismas o aquél constituirían dicho contenido. Se trata de evitar así derivas como las propias del idealismo subjetivo berkeleyano. Una segunda acotación se halla contenida ya en lo antedicho, pero precisa de ulteriores matizaciones habida cuenta de la centralidad de esta idea dentro de este marco teórico. Se trata de la idea según la cual el carácter fenoménico de un estado mental se encuentra exhaustivamente constituido por su contenido intencional.[8] Según este axioma de las teorías representacionales, para cada carácter fenoménico concreto F existe un contenido intencional I tal que un estado con ese carácter F no es otra cosa que un estado mental con I como contenido suyo; esto es, el carácter fenoménico de un estado mental se halla completamente especificado por su contenido representacional. Este axioma implica que para cada carácter fenoménico experienciable existe un contenido intencional específico que lo determina, y que todo estado mental que porte del modo apropiado ese contenido específico tendrá el correspondiente carácter fenoménico. Las teorías representacionales de primer orden coinciden pues en su conceptualización del carácter fenoménico de la experiencia como dependiente del contenido representacional, hallándose todas ellas comprometidas con la tesis según la cual no puede darse variación fenoménica en ausencia de variación representacional: la identidad representacional supone identidad fenoménica, de tal modo que dos experiencias sólo pueden diferir en tanto difieran los contenidos que las determinan. A menudo se ha argumentado contra esta tesis central de las teorías representacionales que existen estados mentales cuyos contenidos intencionales son, cuando menos, difíciles de explicitar, como los dolores o los estados de ánimo. La réplica de los teóricos representacionalistas ha consistido en tratar de definir de diversos modos el contenido de semejantes estados –así, por ejemplo, ya Armstrong (1968), y más recientemente Tye (1995) o Bain (2003), han defendido que las experiencias de dolor representan de diferentes formas distintos tipos de disfunción

[8] Esta exhaustividad representacional trae consigo dos consecuencias. En primer lugar, una clase de superveniencia representacional que adoptaría la siguiente forma (y en la cual se basarían la inmensa mayoría de los intentos antifuncionalistas de desarticular los fundamentos de las teorías representacionales de la conciencia): no puede haber diferencias en la conciencia fenoménica sin diferencias en el contenido representacional. En segundo lugar, la denominada tesis de la *transparencia de la experiencia* (vid., v. g., Crane, 2003; Dretske, 2003; Harman, 1990; Tye, 1992; 1995; 2000 y, especialmente, 2002; para una elocuente discusión de dicha tesis vid. Stoljar, 2004): dado que nuestra conciencia estaría exhaustivamente constitutita por los contenidos de nuestras representaciones mentales, nada más allá del modo en que las cosas son representadas debería aparecer en nuestra experiencia consciente y, así, centrando nuestra atención en ésta, todo lo que habríamos de hallar serían los señalados contenidos de nuestras representaciones.

o daño orgánico–. No obstante, el principal problema de estas teorías representacionales y, de hecho, de las teorías representacionales en general, consiste en ofrecer una teoría de la representación que no resulte conflictiva y se muestre capaz de dar cuenta de los datos empíricos y las objeciones conceptuales. Tal teoría no ha sido lograda, y tanto la proliferación de intentos encaminados a ello como la incapacidad de todos y cada uno de los mismos para sortear los obstáculos que a dicha empresa se oponen, hacen que a nadie extrañen los recientes llamamientos al destierro de la noción de representación de la filosofía de la mente y las ciencias cognitivas. Con todo, las teorías representacionales han seguido el curso de las teorías de la representación, amoldándose en su mayoría al externismo que inundara dicha área desde las propuestas de Kripke (1980), Putnam (1975) y Burge (1979) en filosofía del lenguaje.[9] El así denominado externismo fenoménico que arraiga en las mismas y que ha sido desarrollado por Dretske (1995: cap. 5; 1996), Lycan (1996; 2001) o Tye (1995; 2009: 193 y ss.) no ha dejado de ser criticado por las antiintuivtivas consecuencias que acarrea, habitualmente ilustradas mediante engendros filosóficos como la Tierra invertida de Block (1990; 1996) o el hombre del pantano de Davidson (1987).[10] Este externismo, según el cual la individuación de un estado mental depende de sus relaciones con su entorno, ha sido precisamente por esto denominado *relacional*: un estado mental que representa a una determinada entidad real debe hallarse con ella en la apropiada clase de relación. Aquí arraiga una nueva consecuencia indeseada de este tipo de representacionalismo, dado que la necesidad de concebir a las representaciones mentales como portadoras de valores de verdad implica que los colores, olores y otras cualidades secundarías deban entenderse como objetivamente existentes.

Lo hasta aquí apuntado dice poco acerca de la naturaleza del carácter fenoménico de los estados mentales. Según estas teorías, el mismo consiste, meramente, en contenido representacional.[11] Pero, ¿todo contenido representacional comporta, por el hecho de

[9] Los tres trabajos son citados en orden de aparición, dado que las lecciones que integran el libro de Kripke fueron dictadas en enero de 1970.

[10] El modo en que las referidas consecuencias antiintuitivas del externismo fenoménico son puestas de relieve haciendo uso de este famoso experimento mental es el siguiente: una réplica de un ser humano cualquiera, idéntica a éste molécula a molécula y surgida espontáneamente por alguna extraña suerte de albur (digamos, un rayo que al caer sobre un pantano organiza de la forma apropiada un volumen dado de materia inorgánica tornándola orgánica), no pasaría de ser, desde la perspectiva externalista, un trozo de carne inconsciente, pues dicha réplica carecería de la apropiada historia causal y evolutiva de relaciones con su entorno.

[11] "El aspecto fenoménico de los estados mentales equivale a cierta clase de contenido intencional" (Tye, 1995: 137τ). "La mente no tiene ninguna propiedad especial más allá de sus propiedades representacionales, sumadas a la organización funcional de sus componentes" (Lycan, 1996: 11τ).

serlo, un determinado carácter fenoménico? Las teorías representacionales de primero orden se dividen en este punto entre las que, como la de Thau (2002), se ciñen exclusivamente a la modalidad y el contenido de las representaciones mentales para dar cuenta de la diferencia habida entre representaciones fenoménicamente conscientes y representaciones inconscientes y, por otra parte, las que, como las de Dretske (1995) y Tye (1995; 2000), apelan a propiedades de las representaciones mentales de cara a alcanzar dicho objetivo. En el caso de la teoría de Tye, que es la que ha gozado de mayor predicamento, estas propiedades incluyen, además del carácter abstracto (esto es, la ausencia del requerimiento de la presencia de un objeto concreto que satisfaga el contenido de la representación) y no conceptual (esto es, la ausencia de un concepto al que vincular cada experiencia, la ausencia de cualquier clase de red de significados que establezcan una relación categórica entre representación y experiencia) propio, como indicábamos, de las representaciones que subyacen en estas propuestas al carácter fenoménico de la experiencia consciente, la disposición, aptitud o preparación ("poised")[12] para influir en la economía cognitiva del organismo experienciante causando o provocando otros estados mentales con carácter fenoménico y/o contenido proposicional: los contenidos de la experiencia se ubicarían pues en la periferia de los sistemas cognitivos de alto nivel y se hallarían preparados para influir en la dinámica de éstos. Estas tres propiedades serían en la propuesta de Tye las tres condiciones en las que se basa la diferencia entre un estado representacional fenoménicamente consciente y un estado representacional no dotado de carácter fenoménico, y particularmente la última, que se presenta dentro de esta teoría como la condición realmente diferencial: un estado representacional subpersonal puede ser abstracto y no conceptual y seguir careciendo de carácter fenoménico hasta que no se encuentre en disposición de afectar del modo señalado a la economía cognitiva del organismo. Desde el punto de vista de Tye, pues, el carácter fenoménico de nuestra experiencia consciente equivale a determinada clase de contenido representacional, una clase contenido que, según su planteamiento (vid. Tye, 2008; 2009), a) se halla connaturalmente vinculada con fenómenos neurofisiológicos, pues cada contenido representacional equivale a un patrón concreto de actividad neurofisiológica; b) incluye componentes sensoriales y afectivos; c) representa las características relevantes de los

[12] Incluimos esta retahíla de sinónimos a causa de la dificultad que implica la traducción de este término. Quizá la mejor solución sea la de Martínez (2008), que lo vierte al castellano como "preparado", vocablo que conserva el sentido de "ser apto" o "hallarse listo" o "en disposición" (en este caso, de permitir la aparición de nuevos estados mentales).

objetos de nuestra experiencia; d) es no conceptual[13] y, por último, e) transparente, esto es, que nos apercibimos de forma inmediata de nuestras experiencias y, adicionalmente, que no necesitamos dirigir nuestra atención en uno u otro sentido para hacerlo. La idea de la transparencia de la experiencia, central en la marco teórico elaborado por de Tye, implica así que aquello que experimentamos son las propias características de los objetos de nuestra experiencia, unas características que, cuando lo que experimentamos son objetos externos y no estados orgánicos o anímicos,[14] son públicamente accesibles y pertenecen al objeto externo que sea el caso antes que a nuestra experiencia del mismo.

La proliferación de teorías representacionales de primer orden a que han asistido las dos últimas décadas ha traído consigo una paralela profusión taxonómica pródiga en signaturas barrocas. Consideramos, por nuestra parte, necesario aludir a una sola distinción más de cara a ofrecer una visión de conjunto cabalmente ajustada a la señalada proliferación: aquélla que hallamos entre las teorías que, como la de Tye, equiparan lo fenoménico con determinada clase de contenido representacional dotado de cierta clase de propiedades, y las teorías que, como las de Lycan o Prinz, aluden además a mecanismos funcionales de cara a precisar su noción de representación. Lycan defiende en este sentido que el carácter fenoménico de la experiencia se encuentra determinado por la economía funcional de los mecanismos de procesamiento a los que ésta debe su existencia. Asume, además, que la red relacional que en su planteamiento vincula causalmente las representaciones en que se basa la experiencia consciente con aquello que las mismas representan es perfectamente explicitable en términos funcionales (Lycan, 1996) y psicosemánticos (Lycan, 2001).

*Teorías representacionales de orden superior*

Por su parte, las teorías representacionales de orden superior consideran necesaria pero no suficiente para la emergencia de la conciencia fenoménica la presencia de las representaciones de primer orden en que se basan las teorías que resumíamos en el apartado anterior, dado que, según las mismas, un requisito adicional indispensable para que

[13] Inicialmente, Tye (1995) elaboró una postura híbrida que daba cabida a la posibilidad de que el contenido representacional refiriera en ocasiones a un modo determinado de presentación de un objeto en lugar de a las características relevantes del mismo. No obstante, siguiendo la estela de Peacocke (2001), abandonaría en sucesivas publicaciones el conceptualismo presente en aquella postura híbrida.

[14] Desde la perspectiva de Tye (2008), el carácter fenoménico de las emociones y estados anímicos supervviene a determinada clase de representaciones, una clase de representaciones, en este caso, referidas a sus contenidos a través de experiencias perceptivas o de cierta suerte de pensamientos acerca de las mismas.

un sistema u organismo sea sujeto de experiencia consciente es que el mismo porte un estado representacional que tenga por objeto alguno de aquellos estados representacionales de primer orden, es decir, se hace necesario un estado representacional de segundo orden acerca de un determinado estado representacional de primer orden.

Cabe distinguir dentro de las teorías representacionales de orden superior entre teorías de experiencia de orden superior y teorías de pensamiento de orden superior. Según las primeras, las postuladas representaciones de segundo orden consisten en cuasi-percepciones acaecidas en virtud de alguna suerte de sentido interno dirigido a los estados mentales de primer orden u orden inferior. De este modo, según las teorías de experiencia de orden superior (vid., v. g., Armstrong, 1978b; Lycan, 1987; 1995), las representaciones de segundo orden tienen lugar de forma análoga a la percepción propiciada por los órganos sensoriales, pero, a diferencia de ésta, aquéllas están dirigidas internamente y funcionan como un mecanismo de escáner interno que monitorea los estados mentales de primer orden. Un estado fenoménicamente consciente es definido en este marco teórico, pues, como un estado con contenido intencional no conceptual que es objeto, mediante cierta suerte de sentido interno, de un estado asimismo dotado de un contenido intencional no conceptual, constituido en este caso por el anterior estado intencional. La diferencia entre un estado fenoménicamente consciente y un estado mental inconsciente consistiría así en que el primero sería procesado por cierta clase de sentido interno para producir un estado de orden superior dotado de un contenido intencional no conceptual que constituiría al tiempo el contenido fenoménico de dicho estado, mientras el segundo permanecería fuera del alcance del señalado sentido interno y no pasaría de consistir en un estado intencional de primer orden, dotado acaso de contenido no conceptual, pero no de carácter fenoménico. La actividad del sentido interno sería de este modo la que produciría contenidos perceptuales de segundo orden en virtud de los cuales determinados contenidos intencionales no conceptuales se hallarían a disposición del razonamiento práctico y el control de la conducta del sujeto. Esos contenidos perceptuales de segundo orden serían precisamente los que darían lugar al carácter fenoménico de los estados de primer orden a los que apuntan, los que decidirían, parafraseando a Nagel (1974), qué es como estar en esos estados o experimentarlos. No obstante, como han señalado diversos críticos (v. g. Dretske, 1995; Güzeldere, 1995; Sturgeon, 2000), un importante problema con el que topan las teorías de experiencia de orden superior estriba en su incapacidad para especificar criterios con arreglo a los cuales demarcar la noción de ese sentido interno que postulan. Así, por ejemplo, siendo como es

el mismo postulado en analogía a los sentidos externos, resulta extraño que no exista una experiencia fenoménica concreta asociada a él, como sucede con éstos, del mismo modo que resulta extraño que no parezca dar el mismo, como éstos, lugar a errores o disfunciones.

Las teorías de pensamiento de orden superior, por su parte, presentan a las representaciones de segundo orden necesarias para la emergencia de la conciencia fenoménica como abstracciones antes que como percepciones directas producidas por alguna clase de sentido interno. Desde el punto de vista de estas teorías, lo que se hace necesario para la emergencia de la experiencia consciente es un pensamiento de segundo orden u orden superior acerca de los estados mentales de primer orden u orden inferior, de forma que un estado representacional de primer orden se hace consciente cuando es objeto de un estado representacional abstracto de segundo orden. Un estado mental fenoménicamente consciente sería así un estado mental que causa de forma no inferencial una creencia, habitualmente inconsciente, según la cual el sujeto está teniendo precisamente ese estado mental consciente; esto es, un estado objeto de un pensamiento de segundo orden causado por él de forma no inferencial: un estado intencional que produce un estado intencional de orden superior acerca de sí mismo. Dentro de las teorías de pensamiento de orden superior cabe distinguir entre aquéllas para las cuales un estado mental cuenta como consciente en tanto el mismo es objeto potencial de un estado representacional de orden superior, como la de Carruthers, y aquéllas para las cuales un estado mental puede ser tenido por consciente en tanto el mismo es ya efectivamente objeto de un estado representacional de orden superior, como la de Rosenthal. En la propuesta de éste "el ser consciente de un estado mental se identifica con la posesión de un pensamiento simultáneo de que uno se encuentra en dicho estado (…), pero dado que un estado mental sólo es consciente cuando está acompañado del pensamiento de nivel superior adecuado, se puede explicar nuestro ser consciente de dicho estado mental mediante la hipótesis de que el mismo causa dicho pensamiento de nivel superior" (Rosenthal, 1986: 335-336τ). Por su parte, en el planteamiento de Carruthers, y en las concepciones disposicionalistas en general, un estado fenoménicamente consciente es un estado mental capaz de causar de forma no inferencial pensamientos de orden superior acerca de sí mismo, pero estos pensamientos de orden superior no tienen necesariamente que darse de forma actual, pudiendo presentarse en potencia, con lo cual estaría eludiéndose la proliferación de procesos cognitivos en curso necesarios para dar lugar a la

experiencia consciente. Para Carruthers, en cualquier caso, contarían como fenoménicamente conscientes, exclusivamente, los estados mentales cuyas propiedades pueden ser reconocidas por nuestras capacidades introspectivas, de tal modo que un estado mental fenoménicamente consciente quedaría definido como un estado mental de nivel superior que daría cuenta de las propiedades primarias del estado mental de nivel inferior que tuviera como objeto y, así, la capacidad de reconocer una experiencia consciente como una experiencia fenoménica de una determinada clase sería lo decisivo de cara a ofrecer una explicación de la experiencia consciente. Lo que ni Carruthers ni ningún otro defensor de las teorías de pensamiento de orden superior acaba de explicar de forma convincente es en qué consisten estas capacidades de reconocimiento, y lo que en ningún caso resultaría justificable dentro de este marco teórico sería la efectividad causal de formas no fenoménicas de intencionalidad o estados mentales inconscientes, dado que el mismo se alza sobre una asunción según la cual sólo a los estados mentales como los que podrían ser objeto de las referidas capacidades de reconocimiento podría suponérseles eficacia causal.

Uriah Kriegel y Kenneth Williford, entre otros, han dado recientemente un paso más dentro del ámbito de las teorías representacionales de orden superior al proponer la que denominan *teoría autorrepresentacional de la conciencia*, según la cual un estado mental es fenoménicamente consciente si y sólo si se representa a sí mismo de forma adecuada. (vid. Kriegel & Williford, 2006). Los planteamientos de orden superior vistos anteriormente necesitaban postular la existencia de dos estados mentales discretos, mientras que para las teorías autorrepresentacionales con uno es suficiente: un estado mental fenoménicamente consciente sería pues un estado intencional con un cierto contenido que a su vez se tendría a sí mismo como contenido. La relación entre el contenido intencional de primer orden y el contenido de segundo orden que semejante clase de estado mental habría de tener ha sido planteada de dos maneras: en una versión ambos niveles aparecen relacionados dentro de una misma estructura, compleja pero integrada, en la que una parte de la misma (análoga a un estado de primer orden) es objeto de otra (análoga a un estado de segundo orden), mientras en otras no se hace necesaria esta distinción entre partes dado que uno y el mismo estado adquiriría, como un todo, propiedades de primer y segundo orden en virtud de su actividad dentro del sistema cognitivo en el que tiene lugar.

En general, puede decirse que todas las teorías representacionales de orden superior parten de un supuesto típico en ciencias cognitivas, el supuesto según el cual el nivel de análisis apropiado para el estudio y explicación de cualquier fenómeno mental es el nivel representacional, del cual puede partirse hacia una explicación del fenómeno que sea el caso ensayada en términos de roles causales, papeles funcionales y contenidos intencionales. Un supuesto aparejado al anterior y que igualmente comparten los proponentes de las teorías representacionales de orden superior de consuno con el grueso de científicos cognitivos es el de que cualquier –o prácticamente cualquier– clase de estado mental puede presentarse en formas conscientes o inconscientes, hallándose para los teóricos representacionalistas la diferencia entre unas y otras en que las primeras, pero no las segundas, son objeto de algún tipo de representación de orden superior.

### 3._Teorías neurobiológicas

Las diferentes teorías neurobiológicas de la conciencia que han visto la luz desde la década de los noventa comparten gran cantidad de supuestos, entre los cuales el decisivo es el de que en la investigación y teorización neurobiológica residen las claves que nos permitirán resolver el problema de la conciencia. A este supuesto básico cabe añadir otros puntos de convergencia, entre los cuales destaca la importancia concedida por las distintas teorías neurobiológicas a los bucles de actividad nerviosa recursiva y coordinada. Sin embargo, los detalles de cada una de las propuestas teóricas particulares difieren ostensiblemente, haciéndolo por tanto asimismo los resortes explicativos que destinan a dar cuenta del modo en que el agua neuronal deviene vino fenoménico. Un aspecto que, en una medida u otra, comparten todas las teorías neurobiológicas es su necesidad de apelar a datos y marcos teóricos tomados de la psicología cognitiva, dado que de lo que se trata es de ofrecer una explicación del modo en que la actividad neurofisiológica produce, precisamente, los fenómenos que la psicología experimental ha venido describiendo e integrando en diversos marcos teóricos. En este sentido, se ha llegado a afirmar (vid., v. g., Koudier, 2009: 87) que la mayoría de las teorías neurobiológicas de la conciencia pueden ser concebidas como extensiones de teorías cognitivas preexistentes –una afirmación un tanto exagerada, dado que su contenido puede ser aplicado de forma estricta únicamente a la teoría del espacio de trabajo global neuronal–. Una forma –quizá un tanto caricaturesca– de presentar esta confluencia consiste en señalar que la diferencia entre unas teorías y otras suele estribar en que los diagramas de flechas y

cajas propuestos en unas y otras difieren sólo en que los dibujados por neurocientíficos añaden a la función psicológica la localización anatómica. En cualquier caso, buena parte de la comunidad neurocientífica considera excesivamente especulativa la teorización cognitiva e insiste en la necesidad en que una ciencia de la experiencia consciente se halla de apelar a mecanismos neurofisiológicos descritos en un vocabulario más o menos neutral antes que a supuestos procesos psicológicos descritos en el habitual vocabulario computacional propio de la psicología del procesamiento de información.

Ofreceremos aquí un breve repaso por las más destacadas entre las teorías neurobiológicas de la conciencia y, en la tercera parte, analizaremos pormenorizadamente una propuesta metodológica estrechamente vinculada con la neurobiología de la conciencia: la neurofenomenología.

Las teorías que repasaremos brevemente en este apartado pueden agruparse en dos categorías. Tendríamos, por una parte, aquellas teorías que han tratado de explicar el origen de la experiencia consciente asociada a una modalidad sensorial –generalmente la vista, aunque existen planteamientos centrados en otras modalidades, como el de Walter Freeman, que partiera de estudios electrofisiológicos sobre el olfato en conejos–, como la de Semir Zeki, la de Viktor Lamme, la de David Milner & Melvyn Goodale y la de Francis Crick & Christof Koch, y por otra, aquéllas que se han ocupado del origen de la conciencia desde enfoques neurofisiológicos más abarcadores, como la de Michael Gazzaniga, la de Vilayanur Ramachandran, la de Rodolfo Llinás, la de Stanislas Dehaene, Lionel Naccache & Jean-Pierre Changeux, la de Gerald Edelman y la de Antonio Damasio.

*La teoría de Semir Zeki*

La teoría de Zeki es sin lugar a dudas la más circunscrita o localista entre las suscritas por neurocientíficos. La misma parte de la asunción según la cual, dado que diferentes aspectos de la conciencia visual se relacionan con la actividad de diversas áreas encefálicas, es decir, dado que no son las mismas áreas de la corteza occipital las que se relacionan, por ejemplo, con la visión del color (V4: una lesión en esta área da lugar a acromatopsia) o el movimiento (V5: una lesión en ésta ocasiona acinetopsia), y dado que la percepción consciente de diferentes atributos visuales (como el color o la forma) no es necesariamente simultánea y puede de hecho obedecer a una secuencia jerárquica –datos experimentales avalan que la ubicación de un estímulo es percibida antes que su

color, y éste, a su vez, antes que su movimiento (vid. Pisella et al., 1998)–, la conciencia no es un fenómeno singular y unificado o unitario, sino que existen diferentes conciencias extendidas en el espacio y el tiempo neuronal y asociadas a regiones encefálicas independientes que el británico denomina nodos esenciales. Según Zeki (1993), la actividad en cada nodo esencial puede dar lugar a resultados perceptivos explícitos que no necesitan de un ulterior procesamiento ni de influencias top-down desde áreas corticales superiores para originar una experiencia consciente. Así, cada estado (micro)consciente de color es en esta teoría el correlato de determinada clase de actividad en V4 del mismo modo que cada estado (micro)consciente de movimiento lo es de determinada clase de actividad en V5. La teoría de Zeki parece eludir el denominado *bindig problem*, esto es, el problema de explicar el modo en que los diferentes atributos de un estado consciente relacionados con la actividad de diferentes áreas corticales y subcorticales se integran para dar lugar a una escena unitaria y coherente, pero de hecho Zeki postula que esta unión de los diferentes atributos de un estado consciente no se halla, como en aquel momento defendían Crick y Koch, a la base de la generación de la experiencia consciente, sino que tiene lugar después de producidas las diferentes "microconciencias" de cada uno de los atributos que la integrarían –Zeki funda esta hipótesis en resultados experimentales obtenidos en estudios en los que los sujetos enlazaban incorrectamente el color de un percepto dado en un momento concreto con la dirección del movimiento registrada 100 milisegundos antes (Moutoussis & Zeki, 1997)–. De este modo, Zeki pospone el *binding problem* e hipotetiza que el enlace de los diferentes atributos es un proceso postconsciente.

> We propose that, if any binding occurs to give us our integrated image of the visual world, it must be a binding between microconsciousnesses generated at different nodes. Since any two microconsciousnesses generated at any two nodes can be bound together, perceptual integration is not hierarchical, but parallel and postconscious (Zeki & Bartels, 1999: 225).

Esta integración postconsciente da lugar a lo que Zeki (2003) denomina "macroconciencia".

> I refer to consciousness of a stimulus that is compound, in that it consists of more than one attribute, as a 'macroconsciousness' to distinguish it from consciousness of a single attribute (e. g. colour), which I designate as a 'microconsciousness' (Zeki, 2005: 1169).

Zeki ha desarrollado junto con sus colaboradores una teoría de esta integración de "microconciencias" en "macroconciencias" a la que denominan teoría de la integración multi-etapa. Según esta teoría, cada sistema de procesamiento perceptivo tiene una cierta estructura jerárquica y, sin embargo, no existe una jerarquía perceptiva en la integración de diferentes atributos, dado que la actividad perceptiva explícita de grupos neuronales pertenecientes a un nivel temprano de procesamiento en un sistema perceptivo concreto puede enlazarse con la de grupos neuronales a un nivel tardío en otro o el mismo sistema. De acuerdo con los autores, los nodos activos cuya actividad resulta enlazada en un momento determinado constituyen unidades funcionales, y un nodo puede formar parte de diferentes unidades funcionales, las cuales se constituyen dinámicamente y pueden integrar información de diferentes etapas y diferentes sistemas de procesamiento perceptivo (Zeki & Bartels, 1998; Zeki & Ffytche, 1998).

No obstante, con esta integración y estas "macroconciencias" no acaba la relación de formas de conciencia postuladas por Zeki, dado que también habla de una conciencia unificada que, afirma, es la única conciencia de la que cabe hablar en singular y que sólo es posible dada la capacidad humana para la comunicación y el lenguaje (Zeki, 2007: 580), encontrándose así estrechamente emparentada con la conciencia de acceso. Zeki opone de este modo su teoría a la del espacio de trabajo global neuronal (vid. infra). Esta teoría considera que la actividad de una difusa red de áreas fronto-parietales es un elemento crítico para el origen de la experiencia consciente. En la teoría de Zeki, esas áreas corticales no juegan ningún papel en la génesis de las diferentes "microconciencias" visuales, de forma que, desde su perspectiva, la teoría del espacio de trabajo global neuronal contaría, como mucho, como una explicación de la conciencia de acceso –o unificada, en la nomenclatura de Zeki–, mientras la conciencia fenoménica, que vincula con sus "micro-" y "macroconciencias", y que el grueso de investigadores considera *el* problema de la conciencia, quedaría fuera del alcance de dicha teoría.

*La teoría de Viktor Lamme*

La teoría de Lamme se basa en la distinción de tres etapas jerárquicamente ordenadas de actividad neuronal relacionada con la conciencia visual. La primera de las mismas implica un curso de actividad neuronal que parte de áreas estriadas hacia áreas extraestriadas, parietales y temporales. Lamme habla de un *fast feedforward sweep* de transmisión de información visual desde el córtex visual primario hacia áreas extraes-

triadas y hacia las vías visuales ventral y dorsal. Se trata de un flujo de actividad nerviosa que recorre hacia adelante la jerarquía de procesamiento del sistema visual pudiendo alcanzar la corteza motora y las regiones implicadas en el control de la conducta. Pero este flujo de actividad es concebido por Lamme como un inflexible "arco reflejo cortical" que conduce la información en una sola dirección hacia las áreas motoras interviniendo en conductas automáticas aunque inteligentes y complejas –Lamme llega a hablar en este punto de "cognitive behaviour"– al permitir una rápida aunque básica categorización visual en clases potencialmente relevantes desde el punto de vista conductual (como cara/no-cara o animal/no-animal). Durante esta primera etapa, el estímulo visual que produjera la señalada actividad neuronal no sería experimentado conscientemente, cosa que cambiaría en la segunda etapa, en la cual tendría lugar un bucle local recurrente que llevaría la actividad neuronal de vuelta a las áreas visuales primarias, además de extenderla horizontalmente a otras áreas corticales. Esta interacción recurrente entre áreas visuales primarias y áreas de asociación es la que, en la propuesta de Lamme, da lugar a la experiencia visual consciente (Lamme & Roelfsema, 2000; Lamme, 2006). El origen de la experiencia consciente se ubica así en una clase de actividad nerviosa propiciada por interacciones córtico-corticales recurrentes mediadas por conexiones *feedback*, que devuelven el flujo de actividad a áreas anteriores en la jerarquía de procesamiento, y conexiones horizontales, que conectan cada peldaño de la jerarquía con diversas áreas distantes (Lamme, 2000; 2003).

Lamme reconoce que sabemos mucho menos del funcionamiento de las interacciones córtico-corticales recurrentes que de la clase de procesos neurofisiológicos implicados en la primera etapa. No obstante, existe evidencia que apunta a la necesidad de esta clase de interacciones en procesos perceptivos complejos relacionados con la interpretación perceptiva, como el agrupamiento perceptivo o la segregación figura-fondo. Lamme cita evidencias procedentes de experimentos en los que procesos neurofisiológicos como los que definen su primera etapa inconsciente no aparecen asociados a experiencia consciente, así como otras procedentes de experimentos de enmascaramiento que apoyan la hipótesis de que la ausencia de procesamiento consciente del estímulo enmascarado responde a una supresión de la actividad recurrente producida por el enmascaramiento. Estas evidencias apoyarían la tesis según la cual una forma recurrente de procesamiento es necesaria para la experiencia consciente, pero Lamme está interesado en defender una tesis más fuerte: que una forma recurrente de procesamiento como la que describe es suficiente para la experiencia consciente y que, de hecho, tal forma de

procesamiento *es* la propia experiencia consciente. Con todo –y Lamme es consciente de ello–, su aserto se halla necesitado de apoyo experimental y ulteriores remaches teóricos. Así, por ejemplo, aun cuando Lamme se ocupa sólo del sistema visual, hace un aserto acerca de la conciencia en general, con lo cual no queda claro si podríamos encontrarnos ante una teoría de "microconciencias" sostenidas por diferentes bucles recursivos paralelos o si dichos bucles habrían de unificarse en uno mayor mediante alguna clase de mecanismo que ofreciera respuesta al *binding problem*. Por otra parte, si una forma recurrente de procesamiento es suficiente para la conciencia visual, ¿tendremos bastante con la interacción entre dos neuronas, dos capas, dos áreas o todo el sistema visual? Lamme comprende que su teoría ha de expandirse y refinarse para responder a estas cuestiones, pero hipotetiza una suerte de darwinismo neuronal dinámico en el marco del cual conglomerados de bucles recurrentes crecen, se expanden, compiten y se absorben o anulan, actuando los "ganadores" como atractores en sistemas caóticos. Diferentes núcleos de procesamiento recurrente podrían, según esta hipótesis, existir simultáneamente, pero grupos mayores de neuronas implicadas en interacciones recurrentes pronto impondrían una solución unitaria a las múltiples posibles interpretaciones de una escena visual. Maticemos para terminar que en este darwinismo neuronal avanzado –meramente– como hipótesis tentativa, la lucha no busca siempre la hegemonía y diferentes núcleos pueden así coexistir, siempre y cuando no constituyan soluciones alternativas e incompatibles al mismo problema.

En la tercera etapa propuesta por Lamme tendría lugar un bucle recurrente no ya local, sino extendido, en virtud del cual la actividad neuronal relacionada con el percepto visual que sea el caso alcanza áreas relacionadas con las funciones ejecutivas y el lenguaje. En esta etapa, a la experiencia consciente surgida en la segunda, se sumaría la posibilidad de informar del percepto y aplicar al mismo procesos ejecutivos y atencionales, con lo cual Lamme estaría trazando una clara distinción entre accesibilidad a perceptos conscientes por parte de procesos relacionados con el lenguaje y la atención, por una parte, y la experiencia fenoménica del percepto, por otra. Lamme afirmará así que la conciencia visual puede de hecho ser algo muy diferente de lo que nuestras intuiciones introspectivas parecen sugerir y, asimismo, algo diferente de lo que la comunidad científica y filosófica ha venido suponiendo. En concreto, propone que no hay un nexo que vincule la conciencia fenoménica con la atención, el control ejecutivo, la capacidad para informar ni, necesariamente, con procesos corticales de alto nivel (Lamme, 2000; 2003). Una consecuencia antiintuitiva tanto de esta teoría como de la de Zeki sería que

un sujeto podría ser consciente de un percepto o atributo aunque no pueda informar introspectivamente del mismo.

*La teoría de David Milner & Melvyn Goodale*

La teoría de Milner y Goodale se basa en la distinción entre dos vías –diferenciadas funcional y anatómicamente aunque interrelacionadas– por las que discurren los potenciales de acción de las neuronas implicadas en el procesamiento de la información visual: la vía ventral y la dorsal. Una vez llegados al área V1 de la corteza occipital los impulsos nerviosos que parten de la retina y atraviesan el tálamo, ambas vías trazan rutas neuronales diversas. La ventral proyecta axones desde V1 hacia los temporales inferiores mientras la dorsal lo hace hacia los parietales posteriores. Según los autores, la primera es responsable de la formación de una representación perceptual integrada y consciente del objeto, mientas la segunda lo sería del control de las acciones motoras dirigidas a éste. Asimismo, la primera vía opera con representaciones de características duraderas de los objetos percibidos que sirven al reconocimiento y clasificación de los mismos y podrían estar dotadas de contenido semántico, mientras la segunda lo hace con representaciones visuomotoras on-line, representaciones momentáneas de rasgos de los objetos percibidos que sirven a la interacción conductual con los mismos en tiempo real. De este modo, mientras la vía ventral se halla implicada en la visión para la percepción, esto es, en la categorización del objeto mediante la comparación de su representación con representaciones de atributos perceptuales almacenadas en la memoria, la vía ventral, al tratarse de un sistema on-line ocupado en tiempo real de la ejecución de fugaces rutinas ceñidas a aspectos concretos del aquí y ahora, lo estaría en la visión para la acción y no necesitaría trazar semejante clase de vínculos con posibles representaciones visuomotoras estables almacenadas en la memoria.

Las conclusiones de Milner y Goodale se apoyan en evidencias provenientes de la clínica que apuntan a una doble disociación entre la visión para la percepción y la visión para la acción. En este sentido, pacientes con lesiones que afectan a la vía ventral padecen agnosia y se muestran incapaces de reconocer objetos cotidianos, incluso sencillos, a pesar de lo cual, sus movimientos al relacionarse con ellos resultan totalmente eficaces: no pueden percibir conscientemente los objetos –mostrando incluso dificultades con sus atributos, como la forma o el tamaño– pero sí manipularlos apropiadamente. Complementariamente, pacientes con lesiones en la corteza parietal posterior, desembo-

cadura de la vía dorsal, padecen ataxia óptica: pueden describir verbalmente los objetos que les son presentados, tanto por lo que a su identidad se refiere como por lo que a sus atributos perceptivos toca, pero tienen serias dificultades a lo hora de manipularlos.

Milner y Goodale asumen que las funciones realizadas por la vía dorsal, al estar relacionadas con el trato en tiempo real con el mundo físico, tienen un carácter automático y se relacionan con procesos que acaecen y se desvanecen de forma fugaz, motivo por el cual la actividad de esta vía permanece inconsciente –pues, a pesar de que existe una fenomenología relacionada con la volición, la propiocepción o la motricidad en el flujo de conducta destinado a la manipulación de objetos, no sucede lo mismo con la información visual utilizada en la vía dorsal para el control de segmentos transitorios y automáticos de dicho flujo–, mientras que la actividad en la vía ventral se relaciona con procesos que se prolongan en interacción con representaciones almacenadas en la memoria y con representaciones que podrían ser transferidas a la memoria de trabajo. A través de esta interacción con representaciones almacenadas en la memoria, la actividad de la vía ventral alcanza representaciones dotadas de contenido semántico. El carácter lento y off-line de las funciones realizadas por la vía ventral cuenta, según los autores, como motivo para que la actividad nerviosa acaecida en ella llegue a producir experiencias visuales conscientes. Con todo, el planteamiento de Milner y Goodale, entendemos, ha de ser comprendido como un intento de distinguir agrupaciones funcionales encargadas de la visión para la percepción y la visión para la acción antes que como una teoría del modo en que la actividad nerviosa da lugar a la percepción consciente.

*La teoría de Francis Crick & Christof Koch*

Como es sabido, Francis Crick llega a las neurociencias con un Nobel en su haber por su labor en un área diferente de la biología, en concreto, con el Nobel de 1962 en Fisiología o Medicina –compartido con James D. Watson y el entonces ya fallecido Maurice H. F. Wilkins– por el célebre modelo de la estructura molecular del ADN en forma de doble hélice. Crick resolvió en su juventud un problema central de la biología y, asistido por su colaborador Christof Koch, dedicó los últimos años de su vida a la investigación del problema de la conciencia. Alcanzar un éxito parejo en esta área fronteriza de la biología hubiera merecido un segundo Nobel, pero lo cierto es que incluso a día de hoy nadie parece poder hacerse aún una idea excesivamente precisa del modo en que pudiera obtenerse semejante logro. En cualquier caso, el prestigio y la probidad de

Crick contribuyeron decisivamente al afianzamiento de los *Consciousness Studies* y, asimismo, a diluir los persistentes prejuicios acerca de la cientificidad del área. Uno de los primeros frutos de esta segunda vocación biológica de Crick sería *The Astonishing Hypothesis*, texto que viera la luz en 1994, cuando el problema de la conciencia comenzaba a ganar pujanza y captar la atención de investigadores procedentes de distintas áreas. El libro, auque alguno de sus pasajes parezca apuntar en una dirección más ambiciosa –ya en el prefacio indica el autor su intención de desentrañar el problema de la conciencia *tout court*–, presenta una exploración de las bases neuronales de la experiencia visual. Crick se enfrenta, pues, al problema de la conciencia partiendo de uno de sus sectores: la percepción visual. Su abordaje se funda en una serie de asunciones, entre las que cabe destacar las propias de la estrategia que Searle (2000a) ha denominado *building block* (un enfoque metodológico que incluiría las teorías neurobiológicas comentadas hasta el momento): 1) dado que la conciencia se nos presenta en muchas modalidades distintas, hemos de escoger una entre ellas, explicarla y avanzar así paso a paso hacia un estudio integral del fenómeno de la experiencia consciente; 2) la visión es un puente privilegiado hacia el estudio científico del fenómeno de la experiencia consciente. A estas asunciones de partida se suma la que cabe denominar asunción biologicista, según la cual la psicología cognitiva –Crick tiene en mente los modelos de Jackendoff y Baars– está buscando en una dirección condenada al fracaso, dado que el nivel apropiado para una explicación científica de la conciencia es el biológico y, así, el enfoque funcionalista adoptado por los psicólogos cognitivos puede, cuando más, contribuir a avanzar en esta dirección, pero no explicar el fenómeno de la experiencia consciente de espaldas a las ciencias biológicas.

Crick parte de la conciencia visual con la intención de elaborar una teoría científica de la conciencia. De este modo, después de presentar el marco general del estudio psicológico y biológico de la percepción visual, despacha en la primera parte de *The Astonishing Hypothesis* sin demasiados miramientos la perspectiva gibsoniana –rescatada hoy por partidarios del enfoque dinámico anti-representacionalista en ciencias cognitivas– y critica el excesivo énfasis computacional de Marr –señalando que el cerebro no se toma la molestia de solucionar complicados problemas matemáticos– para mostrar a continuación sus simpatías hacia enfoques como el de Ramachandran, que pone a la biología en el núcleo de la ciencia de la percepción visual. Una vez puestas sobre la mesa las cartas de su linaje teórico, Crick hace constar que entiende perfectamente bien la importancia del nivel teórico en neurociencias: no basta con acumular

datos y describir los fenómenos a los que las nuevas técnicas de investigación nos permiten aproximarnos cada vez con mayor detalle, sino que, para entender cómo funciona el sistema nervioso, será en cualquier caso necesario construir modelos teóricos acerca del modo en que –y los mecanismos por los cuales– unos conjuntos de células influyen en otros dando lugar a los distintos fenómenos y procesos que denominamos "mentales", y tanto los modelos de redes neuronales artificiales como los modelos propuestos por psicólogos cognitivos podrán aportar orientaciones en este sentido, pero, según Crick, la última palabra la tendrá en cualquier caso la neurobiología.

El modelo teórico que Crick y Koch propusieran a principios de los noventa se halla estrechamente relacionado con el mencionado *binding problem*. La imagen que vemos integrada parece ser producto de la actividad nerviosa de diferentes áreas encargadas de discriminar, por separado, forma, orientación, color o movimiento. Nadie ha visto el color de un objeto sin forma o el movimiento de un objeto que no estuviera orientado en una u otra dirección, así que, ¿cómo se produce la integración de estas características? A pesar de significativos avances, este problema es aún hoy, pasado ya un cuarto de siglo desde que Crick y Koch propusieran su modelo, lugar para la discusión y la especulación. El tipo de actividad nerviosa que los autores entienden que se halla a la base de nuestra experiencia consciente unitaria se encuentra vinculado con la sección rítmica de la orquesta nerviosa. Según Crick y Koch, dejando de lado las metáforas, las neuronas de las diversas áreas relacionadas con el procesamiento de cada una de las aludidas características enlazarían éstas entre sí mediante el siguiente mecanismo: disparando sus potenciales de acción conjunta y sincrónicamente, en un ritmo próximo a una oscilación gamma (entre 35 y 75 hertzios), una conclusión que Crick y Koch asientan en resultados obtenidos por el equipo de Wolf Singer y Charles Gray y por el de Reinhard Eckhorn. Crick (1994) especula que la actividad nerviosa que da origen a la conciencia parte de las capas inferiores de la corteza (la quinta y la sexta), que expresarían los resultados locales del procesamiento acaecido en capas superiores. Esta actividad en las capas inferiores no se haría consciente, propone, a menos que esté sostenida por alguna forma de memoria a corto plazo, lo cual probablemente requiera a su vez de un circuito reverberador efectivo desde la capa cortical sexta hasta el tálamo y, viceversa, hasta las capas corticales cuarta y sexta. Crick habla de "unidades de procesamiento", conjuntos de áreas corticales en el mismo nivel de la jerarquía visual que proyectan mutuamente axones hacia la capa cuarta de cada una de ellas. Cada juego de estas áreas corticales está estrechamente conectado a una pequeña región del tálamo, que coordina

las actividades de sus áreas corticales asociadas sincronizando sus disparos. El modelo de Crick implica pues, como la práctica totalidad de teorías neurobiológicas de la conciencia, lo que él denomina circuitos reverberatorios entre el tálamo y el córtex.

A diferencia de las dos propuestas teóricas que presentaremos a continuación, Crick y Koch (Crick & Koch, 1990; 1998) entienden que los mecanismos básicos de la conciencia visual están presentes en mamíferos tanto primates como no primates y son así independientes de procesos cognitivos superiores como el lenguaje.

*La teoría de Michael S. Gazzaniga*

La teoría de la conciencia de Michael S. Gazzaniga se basa ya, a diferencia de las teorías neurobiológicas hasta el momento comentadas, en una concepción global del funcionamiento de la mente y el sistema nervioso. Partiremos hacia la presentación de la misma de unas breves pinceladas introductorias acerca del área en que el de Los Ángeles trabaja, un área a caballo entre la psicología y las neurociencias: la neurociencia cognitiva. Veremos en nuestra concisa introducción a la génesis de la disciplina aparecer a Gazzaniga como uno sus los fundadores y ofreceremos a renglón seguido una panorámica de su concepción del funcionamiento de la mente y el sistema nervioso, de la cual su teoría de la conciencia puede considerarse un mero corolario.

En los textos de neurociencia cognitiva suele aludirse bien a David Marr, bien a Michael S. Gazzaniga y George A. Miller como fundadores o pioneros de la disciplina. Al primero se le atribuye a menudo el mérito de haber creado la base para ulteriores desarrollos en el área dado que su feraz *Vision*, publicado póstumamente en 1982 –dos años después de la prematura muerte de Marr–, ha ejercido una constante influencia. El texto puede ser leído hoy como el primer conato integrador en el que el análisis del procesamiento de la información visual se realiza atendiendo a los niveles *computacional*, *algorítmico* y *de implementación*. La propuesta de Marr se dirigía, pues, a la integración de planteamientos provenientes de la psicología cognitiva, la inteligencia artificial y la neurofisiología. Los textos que, por su parte, aluden a Gazzaniga y Miller relatan el nacimiento de la disciplina haciendo referencia a una curiosa anécdota.[15] Según ella, la primera discusión acerca de la necesidad y el modo de trazar puentes entre la psicología cognitiva y las neurociencias habría tenido lugar a finales de la década de los

[15] Una anécdota, no obstante, que algunos comentaristas presentan como el punto de partida de la disciplina (vid., v. g., Brook & Mandik, 2004).

setenta en el asiento trasero de un taxi neoyorquino. En el curso de esta conversación, Gazzaniga y Miller habrían acuñado e intentado comenzar a dotar de sentido a la locución «neurociencia cognitiva».

La neurociencia cognitiva surge, en resumidas cuentas, como una extensión de los programas teóricos y heurísticos de las neurociencias en la cual éstas confluyen con los métodos y resultados de la psicología cognitiva. En este punto de confluencia, el estudio anatómico y fisiológico del sistema nervioso deviene análisis del procesamiento de información por parte del mismo, en un campo de investigación interdisciplinar cuyo objeto de estudio coincidiría con las que Luria (1962) denominara *funciones corticales superiores* –percepción, atención, memoria, lenguaje, planificación–. Se trata, pues, de una rama de las neurociencias sobre la cual ejercen un significativo influjo los planteamientos desarrollados a partir del Simposio de Hixon (*Cerebral Mechanisms in Behavior*, Pasadena, 1948) y la Conferencia Dartmouth (*Symposium on Information Theory*, MIT, 1956), es decir, de la rama de las neurociencias dedicada al estudio experimental de la cognición desde supuestos a caballo entre la biología y la teoría de la información.

La referida acuñación de Gazzaniga y Miller alcanzaría oficialidad académica por vez primera en un curso de neurociencia cognitiva que ambos ofrecieran bajo tal denominación a comienzos de la década de los ochenta en el Cornell Medical College con la intención de, tal y como Gazzaniga señala en el *Handbook of Cognitive Neuroscience*, publicado en 1984,[16] mostrar la orientación del área a la coordinación de los métodos de estudio del cerebro con los de las demás disciplinas cognitivas. Un segundo paso reseñable hacia la oficialidad sería la aparición en 1989 del primer número del *Journal of Cognitive Neuroscience*, sólo un año después de que Gazzaniga lograra obtener de la Fundación James S. McDonnell y el Charitable Pew Trust un importante programa de financiación para la investigación y formación en neurociencia cognitiva.

Otro hito destacable en este momento de consolidación de la disciplina fue la publicación del artículo "Perspectives in cognitive neuroscience", de 1988, en el que Patricia Churchland y Terrence Sejnowski patrocinan el paradigma neuroconexionista en neurociencia cognitiva, una tendencia contra cuyos excesos, como veremos, alzará después la voz Gazzaniga. En el señalado artículo, Churchland y Sejnowski llaman la atención sobre las restricciones que la neurobiología ha de imponer a los modelos compu-

[16] Este libro, editado por Gazzaniga, aparece –junto a textos como *Cognitive Neuroscience: Developments Towards a Science of Synthesis*, publicado por Posner, Pea y Volpe en 1982, y *Mind and Brain: Dialogues in Cognitive Neuroscience*, editado por LeDoux y Hirst en 1986– como documento fundacional de la disciplina.

tacionales al tiempo que defienden la necesidad de una teorización puramente neurocumputacional, ya que, por contraposición a las referidas restricciones neurobiológicas, dudan que todos los detalles de la cognición vayan a sernos accesibles por precisos que lleguen a ser nuestros conocimientos del nivel neurofisiológico.[17] Churchland y Sejnowski trazan en este artículo una línea que prolongarán cuatro años más tarde en un texto que puede ser leído hoy como la biblia de la tendencia conexionista en neurociencia cognitiva (Churchland & Sejnowski, 1992).

Frente a la orientación empirista radical del grupo neuroconexionista de La Jolla abanderado por Churchland y Sejnowski muestra Gazzaniga algunas reservas de raigambre biológica cuyo tenor trae a las mientes el posterior ataque de Pinker al mito de la plasticidad cerebral ilimitada en el segundo capítulo de *The Blank Slate* (Pinker, 2002). Gazzaniga cita en *The Mind's Past* a Sejnowski y presenta a través de él sus puntos de desacuerdo con el grupo conexionista de La Jolla. En el núcleo de estos puntos de divergencia se halla la polaridad genes/entorno en el desarrollo de la mente. Desde la perspectiva conexionista (Gazzaniga tiene en mente a Sejnowski, Churchland y Grush antes que a Rumelhart, McClelland y Hinton), el contacto con el entorno jugaría un papel fundamental en el desarrollo de los rasgos funcionales de la corteza cerebral mientras la especificad genética desempeñaría un papel secundario. Gazzaniga, por su parte, pretende elaborar un marco en el que la teoría de la evolución ofrezca las claves para comprender el funcionamiento del cerebro, el modo en que el mismo origina nuestra vida mental. Desde su punto de vista, la respuesta rápida a la pregunta acerca del origen de las funciones del sistema nervioso y el porqué de los diferentes aspectos de nuestra vida mental puede enunciarse de forma bien concisa: aptitud biológica, es decir, éxito reproductivo. Desde esta óptica, el sistema nervioso no sería sino una enmarañada red de adaptaciones. Según Gazzaniga, pues, en el curso de la evolución fueron seleccionados distintos mecanismos cerebrales que concibe como sistemas altamente especializados. Así, frente a la perspectiva conexionista, que Gazzaniga equipara con la conductista dada la insistencia de ambas en la influencia determinante del aprendizaje, argumenta que el aprendizaje no crea ni da forma de modo decisivo a nuestras capacidades mentales, sino que los estímulos ambientales topan, en su discurrir por el sistema nervioso central, con capacidades evolutivamente preinstaladas en él. Este punto de desencuentro

[17] Esta idea ha sido interpretada (Canseco, 2007) como una invitación a despertar del sueño de Cajal – prosiguiendo con esta afortunada terminología de inspiración kantiana se nos propone en el artículo citado que Churchland y Sejnowski (1992) invitan a despertar del sueño de Boole y del sueño de Marr, es decir, de la clara delimitación de los niveles computacional y algorítmico.

entre planteamientos por lo demás no demasiado distantes podría ilustrarse como sigue: aprendizaje como guía del desarrollo frente a desarrollo (sometido a un potente control genético) como hilo conductor del aprendizaje. Frente a la tendencia neuroconexionista, Gazzaniga defiende que el cerebro viene ya muy equipado de fábrica (Gazzaniga, 1998: 170 del original, 215 de la traducción), y que se trata, además, de un equipamiento modular.

El modularismo de Gazzaniga atenta contra intuiciones fuertemente arraigadas en nuestra concepción de sentido común de la mente, así como contra la forma consuetudinaria y culturalmente gestada de interpretar al sujeto agente.

> Two thousand years of Western thought has urged the view that our actions are the product of a unitary conscious system (Gazzaniga, 1985: 81 del original, 119 de la traducción).

Frente a esta concepción y estas intuiciones, la concepción modular de Gazzaniga presenta a la mente como constituida por diferentes unidades específicas dedicadas al procesamiento de clases particulares de información. Así, el cerebro –y con él la mente, pues la modularidad de la que habla Gazzaniga no es un concepto exclusivamente psicológico, sino que tiene, según el autor, una auténtica base anatómica (Gazzaniga, 1985: 128 del original, 180 de la traducción)– estaría organizado en sistemas modulares de procesamiento relativamente independientes que operarían en paralelo sobre clases específicas de información.

La noción de intérprete remata la concepción modular de Gazzaniga. Dicha noción apela al modo en que generamos creencias y elaboramos teorías acerca de nuestro comportamiento y el de las cosas y personas que nos rodean. El *intérprete del cerebro izquierdo* sería un módulo entre otros, un componente cerebral especial ubicado en el hemisferio dominante (generalmente el izquierdo en los diestros), en el que se hallan las áreas responsables de la generación (área de Broca) y la comprensión (área de Wernicke) del lenguaje. Dicho componente cerebral especial halla su razón de ser en buscar explicación a lo que le rodea, empezando por los resultados de la actividad del resto del cerebro, plagado de módulos que pueden actuar con independencia dando lugar a conductas a las que el intérprete dotará de sentido aunque en principio, por lo que a él respecta, carezcan por completo del mismo. El intérprete, en resumidas cuentas, trata de "explicar los acontecimientos internos y externos estableciendo relaciones causales que den orden y coherencia a la experiencia. Para ello, construye teorías sobre la realidad y

sobre nosotros mismos a través de historias narrativas que nos proporcionan un sentido de unidad" (Enríquez de Valenzuela, 2014b: 263).

Gazzaniga comenzó su carrera investigadora de la mano de Roger Sperry estudiando a pacientes con el cerebro dividido (vid. Gazzaniga, 2005). Esta expresión, «cerebro dividido», refiere a la peculiar condición anatomofisiológica de pacientes epilépticos sometidos a un tipo especial de cirugía: la comisurotomía, un proceso quirúrgico consistente en escindir ambos hemisferios seccionando el cuerpo calloso para aislar los focos de las crisis. El hemisferio no dominante carece de lenguaje, y lo que quiera que suceda en sus módulos sólo se le muestra al investigador y al propio sujeto conductualmente. En la teoría de Gazzaniga, conciencia y lenguaje se encuentran inextricablemente imbricados: los resultados del procesamiento realizado por módulos situados en el silente hemisferio no dominante pueden manifestarse conductualmente, pero no alcanzar la conciencia del paciente. A lo largo de años de experimentación con pacientes con el cerebro dividido Gazzaniga halló una y otra vez que la información procesada por el hemisferio derecho daba lugar a patrones conductuales (basados en procesamiento de información alejado del alcance del intérprete, aislado en su hemisferio dominante) acerca de los cuales el intérprete fabulaba indefectiblemente desde su parlanchín hemisferio izquierdo. Así, por ejemplo, al mostrar el imperativo «levántate» al hemisferio no dominante –que no puede expresarse lingüísticamente pero sí comprender el lenguaje–, el paciente se levanta. Cuando se le pregunta (al hemisferio dominante, que será el que produzca la respuesta verbal) por qué lo hace, el intérprete no puede menos que buscar coherencia y ofrecer una perspectiva que no eche por tierra la idea de que, por así decir, uno se sienta a sus propios mandos, así que responde: "me levanto porque tengo sed: voy a beber algo".[18]

Diversos módulos dotados de un alto grado de especificidad y desconectados del intérprete pueden dar lugar a comportamientos que sólo tras ser objeto del análisis de éste pasan a formar parte de eso que llamamos yo y, así, de eso que llamamos conciencia, pues, tal y como indica Gazzaniga (1985: 135 del original, 188 de la traducción), ésta es producto de la actividad del intérprete. Gazzaniga ha venido desde los setenta no sólo vinculando su noción de conciencia con la capacidad del sujeto para informar lingüísticamente de sus estados internos, sino utilizando esta capacidad como criterio empírico para delimitar el carácter consciente o inconsciente de una información o estado

[18] Otra clase de desorden de la conciencia a cuya elucidación ha aplicado Gazzaniga la noción de intérprete ha sido la esquizofrenia (vid. Gazzaniga, 1988: cap. IV).

mental (vid. Gazzaniga, LeDoux, & Wilson, 1977). No obstante, recientemente Gazzaniga, aviniéndose al esquema descriptivo y las conclusiones de Damasio en *The Feeling of what Happens* (vid. Gazzaniga, 2008: 279 del original, 289-290 de la traducción; 320 del original, 329 de la traducción), ha aceptado la pertinencia de la atribución de conciencia a animales no humanos y, por tanto, no lingüísticos: concede que la sentiencia, la capacidad de experimentar sensaciones elementales, se halla ampliamente extendida en el mundo animal, pero entiende que sólo entre comillas puede decirse que suceda algo análogo con las formas superiores de conciencia, que no sólo implican una unidad coherente de recuerdo del propio pasado y proyección del futuro, sino asimismo el correcto funcionamiento de capacidades cognitivas superiores como la memoria episódica, la metacognición o la autoconciencia o identidad reflexiva que, según Gazzaniga, resulta extremadamente comprometido afirmar que se encuentren presentes en animales no humanos.

*La teoría de Vilayanur S. Ramachandran*

Al igual que Gazzaniga, y echando mano de un supuesto fundamental de la neuropsicología, Ramachandran entiende que el mal funcionamiento del cerebro puede ofrecernos valiosas pistas para comprender el modo en que da lugar a nuestra vida mental, y en particular a nuestra experiencia consciente.[19] No obstante, en lugar de centrarse, como Gazzaniga, en los cerebros divididos, Ramachandran ha dedicado gran parte de su investigación a alteraciones como la sinestesia, los miembros fantasma, la anososgnosia o la epilepsia. También al igual que Gazzaniga, y a diferencia de los autores de los que nos ocuparemos a continuación, Ramachandran no ha puesto en el punto de mira de su labor investigadora la neurobiología de la conciencia ni ha desarrollado una teoría explícita y detallada al respecto. Sin embargo, al igual que en el caso de Gazzaniga, son muchas las aportaciones y observaciones originales que a lo largo de su prolífica carrera ha ofrecido acerca del modo en que las neurociencias pueden contribuir a una comprensión científica de la conciencia.

[19] La forma en que Ramachandran se ha apoyado en la noción de intérprete de Gazzaniga para la elaboración de su explicación de la anosognosia (vid. Ramachandran & Blakeslee, 1998; Hirstein & Ramachandran, 1997) da cuenta del grado en que ambos comparten no sólo el señalado supuesto fundamental de la neuropsicología, sino, adicionalmente, marcos teóricos concretos acerca, por ejemplo, del papel que el lenguaje desempeña en nuestra economía mental.

Ramachandran entiende que no existe una brecha insalvable entre nuestra experiencia consciente y sus bases neurofisiológicas como la aireada por gran cantidad de filósofos desde Nagel, sino que, en una vena wittgensteiniana, propone que la misma es sólo una apariencia derivada de nuestros usos lingüísticos (Ramachandran & Blakeslee, 1998: 231; Ramachandran & Hirstein, 1997: 432), aunque admite que es aún necesaria gran cantidad trabajo teórico y experimental para demostrar que, en efecto, esa insalvable brecha entre fisiología y fenomenología nunca existió. Uno de los aspectos fundamentales de su perspectiva, que comparte también con la de Gazzaniga, es que la misma concede un papel destacado al lenguaje en su intento de determinar el origen de la experiencia consciente. Para Ramachandran, la tarea central del cerebro es la de elaborar y manipular representaciones del entorno y del cuerpo del organismo. De este modo, los diferentes sistemas neurofisiológicos encargados de elaborar las aferencias sensoriales y propioceptivas se hallan inmersos en la labor de operar con representaciones del cuerpo y el entorno. En una etapa evolutiva previa a la aparición de *Homo sapiens*, propone Ramachandran, en el linaje de los primates evolucionó la capacidad de integrar esas representaciones con memorias sensoriales y surgió la posibilidad de manipular mentalmente símbolos. Esa capacidad habría llegado con *Homo sapiens* a un grado de refinación gracias al cual se habría abierto la posibilidad de operar con representaciones de representaciones (o metarrepresentaciones), siendo ése el momento de la historia evolutiva en que la conciencia tal y como la conocemos habría irrumpido en el planeta Tierra. Así las cosas, puede hablarse de la teoría de Ramachandran como asentada en un basamento neurocientífico pero estrechamente vinculada con las teorías representacionales de orden superior, dado que, según la misma, la existencia de la conciencia se debe a la evolución de una capacidad representativa refinada hasta el extremo de poder operar con representaciones de representaciones, y el uso de estas metarrepresentaciones se apoya en el carácter lingüístico de la mente humana –dicha capacidad metarrepresentativa aparece en este marco teórico como facilitadora del lenguaje y, al tiempo, como facilitada por el mismo (Ramachandran, 2004a: 103)–, asentado a su vez en el carácter social de nuestra especie. A pesar de defender que la conciencia se alza sobre una base representacional tan compleja, Ramachandran ha vinculado su noción de *qualia* antes con procesos perceptivos que con procesos cognitivos, llegando a atribuir a las creencias y la imaginería mental una sensación subjetiva a la que cabe denominar *qualia* sólo en un sentido atenuado: "beliefs lack strong qualia" (Ramachandran & Hirstein, 1997: 440). Su noción de *qualia* pretende, por otra parte, salvar los abismos misterianos que

tan habitualmente aparecen como trasfondo de dicha noción al afirmar que existen buenas razones evolutivas para la existencia de los *qualia*, en particular, que ellos juegan un papel primordial en la facilitación de la conducta no automática basada en la toma de decisiones. No obstante, y a pesar de vincular su noción de *qualia* antes con la percepción que con operaciones mentales como la manipulación de imágenes mentales, el recuerdo o el pensamiento, Ramachandran propone que para experimentar aquello a lo que denominamos *qualia* se hace necesaria una mente capaz de manipular metarrepresentaciones (vid. Blackmore, 2006: 186-197). Adicionalmente, apunta a la implicación de la amígdala y otras estructuras del lóbulo temporal en la experiencia consciente –"one needs the amygdala and other parts of the temporal lobes for seeing the *significance* of things to the organism" (Ramachandran & Hirstein, 1997: 450)–[20] y delinea en este punto un significativo paralelismo entre su teoría y la de Jackendoff, según la cual, como vimos, la conciencia se asocia con representaciones intermedias en lugar de con las primeras o las últimas etapas del procesamiento perceptivo.

Las comentadas hasta aquí, o bien eran teorías de la experiencia consciente asociadas a una sola modalidad sensorial, o bien especulativas y deslavazadas apreciaciones expuestas como corolarios de la concepción general de la mente de destacados neurocientíficos. En lo sucesivo comentaremos cuatro teorías neurobiológicas globales y específicas, esto es, no formuladas para una sola modalidad consciente ni como sugerencias o consecuencias que cupiera extraer de marcos teóricos elaborados con finalidades diferentes de la de ofrecer una explicación científica de la conciencia.

*La teoría de Rodolfo Llinás*

La teoría de Llinás tiene a la base su noción de mente –que prefiere denominar "mindness state" (Llinás, 1987)–. Con la misma se refiere el colombiano a la clase de estados funcionales del cerebro que producen imágenes sensomotoras. Según Llinás, dicha clase de estados neurofisiológicos habría surgido en la filogénesis dada su capacidad para la predicción. Los organismos dotados de sistema nervioso se mueven, y la capacidad de predecir las consecuencias del movimiento supone una evidente ventaja evolutiva. En este sentido, en la capacidad para moverse de los animales estribaría la

[20] Cursivas en el original.

evolución del sistema nervioso, y la evolución de la mente habría seguido así un curso que Llinás define como una internalización del movimiento. Desde este punto de vista, el control neurofisiológico del movimiento sería, por así decir, el culpable de la evolución de la mente (Llinás, 2001: 50). Llinás se basa en estudios que realizara a mediados de los setenta acerca del control motor por parte del núcleo olivar inferior y, asimismo, en estudios realizados por Wolf Singer y Charlie Gray mostrando la existencia de actividad sincrónica ampliamente distribuida en la corteza cerebral mamífera para defender la centralidad de los mecanismos de resonancia oscilatoria en la neurofisiología de la cognición. La internalización de la motricidad depende así en su propuesta de la sincronización de la actividad electrofisiológica producida por mecanismos de resonancia oscilatoria. La misma clase de procesos y mecanismos se hallan, según Llinás, a la base de la experiencia consciente, que no sería sino el producto de la sincronización de determinados cursos de actividad neurofisiolíogica en el sistema talamocortical.

Llinás apoya sus conclusiones en resultados experimentales que obtuviera durante la primera mitad de los noventa y que apuntan a la existencia de una clase de oscilación talamocortical coherente de 40 hertzios –nuevamente una actividad gamma– que desaparece durante el sueño profundo pero nos acompaña tanto durante la fase REM –fisiológicamente similar a la vigilia y poblada de esa forma de experiencia consciente que denominamos ensoñaciones– como durante la vigilia.[21] Además, y de acuerdo con la diferente economía de relaciones con el entorno que mantienen la conciencia onírica y la de vigilia (Llinás & Paré, 1991), la señalada coordinación neuronal orquestada en oscilaciones de 40 hertzios se ve afectada y se reorganiza ante estímulos externos durante la vigilia, pero no durante la fase REM (Llinás & Ribary, 1993). Llinás propone que esta actividad gamma refleja las propiedades resonantes del sistema talamocortical, que estaría en sí mismo dotado de una intrínseca actividad oscilatoria de 40 hertzios (Llinás, 1990; 2001).

Otro extremo reseñable de la teoría de Llinás parte igualmente de consideraciones acerca de la motricidad. Llinás denomina *patrones de acción fijados* (FAPs, por sus siglas en inglés) a conjuntos altamente automatizados de patrones motores directamente disponibles para la ejecución de movimientos coordinados y bien definidos (como ca-

[21] "In electrophysiological terms, waking and paradoxical sleep are fundamentally identical states" (Paré & Llinás, 1995: 1155). No obstante, Hobson, entre otros, ha subrayado en repetidas ocasiones las diferencias entre ambos estados, principalmente desde el punto de vista neuroquímico, pero también desde un punto de vista fisiológico más amplio y, asimismo, desde el punto de vista fenomenológico (vid., v. g., Hobson, 2001; Hobson & Fristonb, 2012: 84; Kahn, Pace-Schott & Hobson, 1997).

minar o tocar la guitarra). Estos patrones, coordinados por los ganglios basales, exoneran al sistema talamocortical de una considerable carga de trabajo. Llinás considera a las emociones como una clase premotora de FAPs que proporcionan el contexto interno de la conducta y, en una línea similar a la posteriormente explorada por Derek Denton (vid. Denton, 2005; Denton et al., 2009), sugiere que la conciencia y la cognición probablemente evolucionaran a partir de estados emocionales que activan FAPs (Llinás, 2001: 168). Esta especulación crea el contexto para la argumentación de Llinás respecto de los *qualia* (Ibíd.: cap. 10). Llinás utiliza el término en sentido amplio para referirse a cualquier clase de experiencia subjetiva, defiende la identidad de los *qualia* con la actividad electrofisiológica de circuitos neuronales y ataca el epifenomenalismo arguyendo que la evolución habría seleccionado la capacidad del sistema nervioso para generar *qualia* dada la refinación de la conducta que ellos posibilitarían –se refiere en este punto, de forma un tanto vaga, a una mejorada capacidad para la elección y la decisión.

*La teoría de Stanislas Dehaene, Lionel Naccache y Jean-Pierre Changeux*

Apoyándose en el modo en que Bernard Baars integrara un amplio cuerpo de datos experimentales procedentes de la psicología y las neurociencias en su *global workspace theory*, Dehaene, Naccache y Changeux desarrollan una implementación neurobiológica de la arquitectura funcional de la teoría del espacio de trabajo global: la *neuronal global workspace theory* (vid. Dehaene & Naccache, 2001; Dehaene et al., 2006). La teoría de Baars sostiene que, ante un insumo sensorial, por ejemplo, existen diferentes cursos de procesamiento de información que compiten y cooperan en la elaboración de hipótesis acerca de la identidad del señalado insumo, hipótesis que confluirían en un espacio común de memoria, el espacio de trabajo global. Este enfoque resultó muy provechoso en la década de los setenta cuando se aplicó al complejo problema que para la inteligencia artificial del momento suponía el reconocimiento del discurso hablado en un ambiente ruidoso. Desde entonces, arquitecturas computacionales basadas en los planteamientos de Baars han sido ampliamente utilizadas en robótica e inteligencia artificial, tanto en programas de cuño simbolista como conexionista. La teoría de Dehaene, Naccache y Changeux ofrece una implementación neurobiológica de la arquitectura de un espacio de trabajo global como el hipotetizado por Baars. Esta implementación de la teoría del espacio de trabajo global postula la existencia de dos clases de elementos: una red difusa de módulos de dominio específico y un espacio cortical global de trabajo. Los

referidos módulos de dominio específico estarían integrados por estructuras corticales y subcorticales encargadas del procesamiento de tipos particulares de información (v. g., colores en el área V4 de la corteza occipital, caras en el giro fusiforme). Cada módulo se halla especializado en el procesamiento de determinado tipo de información y en la realización de tareas concretas, y en este sentido son totalmente heterogéneos, a pesar de lo cual comparten una serie de características que permiten tratarlos como entidades análogas. Destaca entre estas características comunes, por una parte, el hecho de que todos y cada uno de estos módulos funcionan de forma inconsciente y automática y, por otra, el de que se hallan encapsulados, es decir, que sus computaciones internas no se encuentran a disposición de las operaciones de otros módulos –aunque estos módulos se presenten como automáticos y encapsulados en un sentido menos fuerte que el fodoriano–. Por su parte, las neuronas del espacio de trabajo global se ubican en el córtex, principalmente en áreas prefrontales, parietales y cinguladas. Dichas neuronas se caracterizan por proyectar y recibir axones córtico-corticales que atraviesan grandes regiones de la corteza cerebral. Se trata de neuronas excitatorias cuyos axones de largo trazado conforman proyecciones córtico-corticales que dan lugar a patrones de activación globales dentro del marco del espacio de trabajo global neuronal que ellas mismas integran, rompiendo así la modularidad de los procesadores de dominio específico y permitiendo un intercambio flexible de información entre los mismos. La teoría del espacio de trabajo global neuronal asume que las señaladas neuronas reciben insumos de dos sistemas neuromoduladores distintos. Por una parte, la intensidad de su espontánea actividad constantemente fluctuante es modulada por los insumos que reciben desde núcleos colinérgicos, noradrenérgicos y serotoninérgicos que producen variaciones en los niveles de arousal. Por otra, el sistema límbico contribuye también a la modulación de la actividad del espacio neuronal global mediante insumos dopamirérgicos directos e insumos indirectos a través de la corteza cingulada anterior y la corteza orbitofrontal.

El espacio de trabajo global que Dehaene, Naccache y Changeux conjeturan es un sistema neuronal distribuido con una conectividad de larga distancia que vincula entre sí módulos encargados del procesamiento de diversas clases de información. Cualquier información manipulada por cualquier módulo que no esté dotado de conexiones directas con las neuronas del espacio de trajo global permanecerá inconsciente. Asimismo, cualquier patrón global de activación puede inhibir patrones alternativos dentro del espacio de trabajo global neuronal, evitando el procesamiento consciente de la información que dicho patrón pudiera portar. Así, esta teoría postula que existe una competición

entre coaliciones nerviosas integradas por neuronas tanto frontales como sensoriales, coaliciones que simultáneamente se ocupan de los mismos fenómenos internos o ambientales. Una estimulación sensorial, por ejemplo, provocaría diferentes patrones de actividad neuronal en la parte posterior del encéfalo, diferentes patrones que competirían entre sí para formar coaliciones dominantes, resultando que algunas de estas coaliciones desencadenan reverberaciones centrales a través de conexiones de largo alcance con los lóbulos frontales, que propiciarían así procesos que permitirían mantener tanto la activación central como la periférica. Una distinción importante para comprender la teoría de Dehaene, Naccache y Changeux es la que cabe trazar entre proveedores y consumidores de representaciones: mientras los sistemas perceptivos proveen representaciones, los mecanismos neuronales encargados del razonamiento, la decisión o el recuerdo las consumen al tiempo que producen otras que son a su vez consumidas por esos mismos mecanismos. Una vez que las representaciones perceptivas provistas por los mecanismos perceptuales son globalmente transmitidas en el córtex frontal, las mismas se hallan a disposición de todo el aparato cognitivo.

Para que un objeto cuente como contenido de un estado mental consciente no es suficiente con que la información relativa al mismo y procesada en distintos módulos sea derivada como insumo a la dinámica de la red neuronal del espacio de trabajo global, sino que, además, a) el contenido de dicho estado mental debe estar representado como un patrón explicito de actividad neuronal en un grupo de neuronas que compendien las características del objeto, b) las neuronas del espacio de trabajo global activas deben poseer un número suficiente de conexiones recíprocas, particularmente con regiones prefrontales, parietales y cinguladas, y, finalmente, c) el mecanismo de amplificación top-down que organiza la dinámica de las conexiones a larga distancia del espacio de trabajo global ha de mantener accesible, definido e intensificado el contenido, pues, en cada momento, las neuronas del espacio de trabajo global sólo pueden manejar una única representación global: la única entre las alternativas posibles elevada a la conciencia en virtud de su contar como objeto del señalado mecanismo de amplificación top-down (Dehaene & Changeux, 2004). En este sentido, una suerte de atención top-down se presenta como condición necesaria para la conciencia en esta teoría, lo cual podría representar un problema para la misma habida cuenta de la dificultad que entraña la definición independiente de ambas nociones, relacionada con las discusiones actualmente

abiertas acerca de la posibilidad de hablar de conciencia sin atención y, alternativamente, de atención sin conciencia.[22]

Según este modelo, los grupos neuronales a los que aludíamos como compendiando las características de un objeto dado, son altamente variables y su actividad puede verse o no reforzada por la estimulación proveniente del medio en el que el organismo se desenvuelve. Este refuerzo puede ser comprendido como inserto en un proceso que deriva en la representación de los estímulos ambientales, lo cual no significa que una red neuronal en particular represente exclusivamente un conjunto concreto de estímulos. Dehaene y Changeux hablan de la existencia de tres clases de representaciones: perceptos; imágenes, conceptos e intenciones; y pre-representaciones. Los perceptos consisten en actividad neuronal determinada por el entorno del organismo y articulada con arreglo a la historia de su relación con dicho entorno. Se trata, así, de una actividad que se desvanece en el mismo momento en que desaparece la estimulación correlativa. Por su parte, las imágenes, conceptos e intenciones son definidos, en una línea que trae a las mientes el presente recordado de Edelman (vid. infra), como objetos mnemónicos resultantes de la activación de un engrama más o menos estable. Por su parte, las pre-representaciones nos son presentadas como patrones inestables de actividad neuronal, patrones transitorios que pueden resultar seleccionados o desechados. Ante cada nueva situación, opera una suerte de proceso selectivo sobre una abundante proliferación espontánea de pre-representaciones, entre las cuales contarían con mayores posibilidades de éxito aquellas que ofrezcan mejores perspectivas para la adaptación a nuevas circunstancias y se ajusten al tiempo a perceptos y conceptos previos.

Desde el punto de vista que este marco teórico ofrece, y de acuerdo con nuestra experiencia cotidiana, la frontera entre lo consciente y lo automático se muestra flexible y dependiente de la práctica. Cuando aprendemos a conducir prestamos una cuidada atención a nuestros movimientos, pero con la práctica los mismos pueden ser realizados sin ningún esfuerzo consciente. La explicación que de este fenómeno ofrece la teoría del espacio de trabajo global neuronal supone que toda vez que un conjunto de procesadores empiezan a comunicarse a través de las conexiones del espacio de trabajo global, este flujo de procesamiento multidominio se convierte, mediante la práctica continuada,

[22] Para una defensa de la independencia y la disociabilidad conceptual y empírica de ambas nociones, vid. Koch & Tsuchiya (2006). Una postura intermedia entre la de Koch y la de Changeux es la de Baars. Aunque, con Koch, presente a la atención y la conciencia como experimentalmente disociables, entiende que de ordinario operan coordinadamente y las concibe como dos aspectos de un mismo proceso en el que la atención actúa como mecanismo de selección de información y la conciencia como mecanismo de integración de la información.

en una forma de procesamiento progresivamente más y más automática que, finalmente, con el trazado de conexiones directas entre los módulos pertinentes, dejará de comprometer a las neuronas del espacio de trabajo global y, por lo tanto, de acarrear actividad mental consciente.

De entre todas las teorías neurobiológicas de la conciencia de las que tenemos noticia, la del espacio de trabajo global neuronal es la única que no hace uso de la distinción entre conciencia de acceso y conciencia fenoménica.[23] De hecho, la teoría de Dehaene, Naccache y Changeux se basa en la asunción de que dicha distinción no es pertinente, mientras el resto de las teorías neurobiológicas presentan a la conciencia de acceso como relacionada con procesos ulteriores y de más alto nivel que los mecanismos neuronales de la necesariamente previa conciencia fenoménica. Los defensores de la teoría del espacio de trabajo global neuronal han llegado a poner en duda la realidad psicológica de esta distinción proponiendo que la capacidad del sujeto para informar de una experiencia consciente es el único criterio fiable de la presencia de actividad consciente, y que los aparentes casos de fenomenología sin conciencia de acceso pueden ser entendidos como casos de la denominada "illusion of seeing" (O'Regan & Noe, 2001), ejemplificada en la ceguera al cambio,[24] que pone de manifiesto nuestra tendencia a pensar que vemos más de lo que en realidad vemos.

*La teoría de Gerald M. Edelman*

Sumándonos al juicio de, entre otros, Baars (2010: 23) o Searle (1997a: 39 del original; 45 de la traducción), entendemos que Gerald Edelman destaca como uno de los neurocientíficos cuyas aportaciones al estudio científico de la conciencia resultan más relevantes por cuanto las mismas se hallan integradas en, quizá, la teoría global más elaborada y sistemática entre las actualmente disponibles en el mercado de los *Consciousness Studies*. Nos detendremos por ello en una exposición de dicha teoría un poco más detallada que las precedentes.

El feraz marco teórico que Edelman elaborara a lo largo de más de tres décadas se asienta en su teoría de la selección de grupos neuronales (Edelman, 1987), una propues-

[23] Puede que de la de Gazzaniga cupiera decir algo análogo, pero, como señalábamos, la suya es una teoría acerca del funcionamiento de la mente en general antes que una teoría neurobiológica de la conciencia propiamente dicha.
[24] Para una concisa introducción a esta noción vid. Wan, Ambinder & Simons (2009).

ta teórica a la que aludiéramos de pasada más arriba y que Edelman desarrolla en línea con planteamientos como los de Changeux (1983) con el propósito de devolver la mente a la naturaleza haciendo hincapié en que un sistema nervioso maduro es producto de un doble proceso selectivo, uno filogenético y otro ontogenético. Según la teoría de la selección de grupos neuronales, el desarrollo ontogenético del sistema nervioso sigue una pauta paralela a la evolución de las especies: la selección rige ambos procesos. El sistema nervioso es, pues, fruto de la selección natural: sus estructuras y funciones responden a ventajas adaptativas. La teoría de la selección de grupos neuronales defiende que, además, el mismo es fruto de la selección neuronal. Esta teoría sostiene que durante el desarrollo ontogenético del sistema nervioso resultan seleccionados aquellos elementos que ofrecen mejores perspectivas adaptativas. Así, Edelman propone en *Neural darwinism* como requisitos para una teoría de la selección neuronal, por una parte, que debe hacerse posible el encuentro con un medio no categorizado y, por otra, que deben poderse amplificar diferencialmente aquellas variantes de mayor valor adaptativo –y esto ha de ir acompañado de una suerte de memoria que posibilite la herencia–. En este sentido, el patrón general de interconexiones neuronales no vendría determinado genéticamente de un modo rígido, aunque el código genético constriñe en cierta medida el proceso selectivo, una constricción que no implica que individuos idénticos genéticamente acaben por desarrollar idénticos diagramas de conexiones, ya que un proceso de selección epigénica –regulado por diversos eventos mecano-químicos acaecidos en el nivel de la célula y el del contexto molecular de su división, crecimiento, translación y diferenciación– tiene lugar en un plano individual: el cerebro del embrión está, por así decir, débilmente determinado por su herencia genética. De este modo, el código genético sería por sí solo insuficiente para especificar la organización acabada del sistema nervioso, de forma que ésta termina por adoptar una configuración determinada partiendo de la base de una excesiva población neuronal que con el transcurso del tiempo y el influjo de la experiencia se ve reducida y perfilada. Es aquí donde opera la selección de las interconexiones y grupos neuronales que resultan más adecuados para el pergeño de la conducta adaptada. Durante la etapa embrionaria se producen las unidades de selección: los grupos neuronales, constituidos como todos funcionales en un proceso de fijación que liga entre sí diversas constelaciones de células. En la fase embrionaria es de capital importancia el papel de las moléculas de adhesión celular, que guían a las células en su crecimiento contribuyendo en la conformación de grupos interconectados de neuronas. Hasta aquí, el producto del proceso selectivo es un repertorio primario de grupos

de neuronas gestado en la etapa embrionaria.[25] Un segundo proceso de selección tiene lugar en el contexto de la primera conducta del recién nacido. En él se modifican los patrones de conexión entre los diferentes grupos neuronales, pero también dentro de cada uno de ellos. Las conexiones entre combinaciones de grupos neuronales se ven reforzadas o debilitadas en el desarrollo de la conducta adaptativa y así, del indicado repertorio primario surge un repertorio secundario en el curso del establecimiento de estructuras de grupos neuronales implicadas en la coordinación de la conducta adaptada, en un proceso en el cual los grupos menos funcionales desaparecen en el marco de una lucha neuronal por la supervivencia: sólo los grupos y las conexiones neuronales más eficaces para la conducta adaptativa –que incluiría tanto procesos motores como perceptivos– estarían en condiciones de seguir luchando. En resumen, se trata de grupos neuronales destinados a desarrollar la misma función y sometidos a principios de selección. En respuesta a críticas como las de Crick (1989), Edelman subraya que los paralelismos metafóricos y detalles analógicos referidos a la teoría de la selección natural pueden ser considerados por la comunidad neurocientífica como prácticamente irrelevantes. Lo verdaderamente significativo sería, antes bien, la concepción del desarrollo y funcionamiento del sistema nervioso a la que su teoría de la selección de grupos neuronales apunta y, asimismo, su potencial explicativo.

El objetivo del darwinismo neuronal es el de presentar el desarrollo y la dinámica del sistema nervioso como guiados no por conjuntos de instrucciones del tipo de los que pueden hallarse en sistemas computacionales, sino por un proceso selectivo.[26] Como sugeríamos, Edelman aplica su teoría de la selección de grupos neuronales tanto al desarrollo del sistema nervioso como a su funcionamiento. Los procesos selectivos posembrionarios postulados en dicha teoría son orquestados por los sistemas de valor aportados por áreas primitivas del cerebro –como los núcleos del rafe, el locus coeruleus, los sistemas colinérgicos y dopaminérgicos o diversos grupos hipotalámicos–, que reflejan el carácter positivo o negativo de una acción o evento para la conducta adaptada del organismo, y en este sentido cabe decir que estos sistemas de valor juegan en la teoría de la selección de grupos neuronales un papel similar a la presión selectiva en la teoría de la selección natural, correspondiendo la magnitud y valencia de las señales de valor a la adaptación del organismo a su medio en la teoría de la evolución.

[25] En Edelman (1988) se esclarecen los fundamentos embriológicos del darwinismo neuronal.
[26] "Digital computers are a false analogue to the brain" (Edelman, 1992: 225).

La teoría de la selección de grupos neuronales conforma la base de la neurofisiología de Edelman, pero el ingrediente fundamental de la misma es el factor de coordinación espacio-temporal de la actividad neuronal producto del proceso de reentrada, un intercambio recursivo de señales entre diferentes áreas encefálicas a través de conexiones recíprocas masivamente paralelas. Esta coordinación espacio-temporal es decisiva para la formación de mapas neuronales interconectados recíprocamente y referidos a estados del medio interno y externo. Edelman ilustra esta noción de mapeado recurriendo a la percepción visual. Con todo, la misma es perfectamente extensible a las diferentes funciones psíquicas humanas.

Haciendo uso de las ideas hasta aquí bosquejadas puede presentarse el esqueleto de la neurobiología de la conciencia de Edelman, pues en su propuesta la base de la experiencia consciente se encuentra en la conexión recíproca y coordinada entre mapas neuronales, una base en forma de red coordinada temporalmente y extendida espacialmente por diversas áreas del cerebro. La coordinación espaciotemporal propiciada por las conexiones masivas de reentrada entre distintos grupos de grupos neuronales se hallarían así a la base de la configuración de una escena coherente y unificada en la que se integrarían gran cantidad de procesos perceptivos, emocionales, mnemónicos y motores. Según la teoría de Edelman, los sistemas neuronales que subyacen a la experiencia consciente surgen en la historia evolutiva como resultado de su capacidad para integrar gran cantidad de insumos sensoriales y respuestas motoras que sobrevienen paralelamente, una integración que permite la discriminación de señales senso-motoras necesaria para la conducta adaptada.

La memoria juega un destacado papel en este marco teórico, hallándose la misma estrechamente relacionada con el sistema de valor aportado por regiones primitivas del cerebro. Este papel crucial de la memoria responde al hecho de que la categorización y el aprendizaje, guiados por los apuntados principios selectivos, se erigen en el planteamiento de Edelman como claves del origen tanto de la conciencia primaria como de la secundaria. La conciencia primaria es lo que Edelman (1989; 1992; Edelman & Tononi, 2000) denomina un presente recordado, el cual concibe como una escena multimodal de eventos sensoriales y motores surgida del acoplamiento del pasado perceptual-categorial y el presente senso-motor con señales procedentes del profundo sistema de valor, un acoplamiento mediante el cual emerge una escena de correlaciones producida por el funcionamiento de vías de reentrada entre diferentes sistemas talamocorticales. Se trataría de una escena en pasado valoral-categorial y presente senso-motor, efecto de la acti-

vación coordinada de redes de reentrada de registros de memoria categorial y reacciones senso-motoras. El animal con este tipo de conciencia poseería un fugaz presente extendido con arreglo al complejo categorial-valoral hacia un pasado no vivido reflexivamente como pasado, sino en tanto que reactualización de memorias ligadas a valores biológicos. Con la aparición evolutiva de la conciencia primaria, que según Edelman habría sucedido dos veces a lo largo de la filogénesis –una con el tránsito de reptiles a aves y otra con el tránsito de aquéllos a los mamíferos–, hace acto de presencia una selectividad discriminatoria radicalmente ampliada, así como una flexibilidad comportamental y una capacidad para la planificación de la conducta sin precedentes. Según Edelman, estas ventajas adaptativas se verían redobladas gracias a los cambios anatómicos (fronto-temporales) en virtud de los cuales accedemos –al menos– algunos primates a la conciencia secundaria, de la cual ya es sujeto un sujeto reflexivo capaz de verse como sujeto de su pasado y su futuro, un sujeto estable capaz de proyectar escenas sensoriales multimodales en ambas dimensiones temporales. Esta conciencia secundaria se halla en el planteamiento de Edelman vinculada al lenguaje, presentándose como condición de posibilidad de su emergencia (vid. Edelman & Tononi, 2000; Edelman, 2004).

La hipótesis del núcleo dinámico supone la aportación más relevante y específica de Edelman y sus colaboradores al área de los *Consciousness Studies* (Edelman & Tononi, 2000). Dicha hipótesis es planteada como una explicación biológica de la conciencia basada en el aspecto informacional de la misma. De acuerdo con dicha hipótesis, la característica crucial de la conciencia es que cada estado consciente constituye una discriminación altamente informativa en cuanto es al mismo tiempo integrada –se experimenta como un todo, de una sola pieza– y diferenciada –al ser diferente de cualquier otro estado consciente entre los posibles en cada momento–. Más allá del aparato matemático que Edelman y colaboradores emplean para formular su hipótesis del núcleo dinámico, cabe señalar que ésta ha de ser concebida como el punto de desembocadura de la selección de grupos neuronales y el proceso de reentrada. El núcleo dinámico correspondería a la activación colectiva, dinámica y coordinada, en el plano de los milisegundos, de grupos neuronales concretos en procesos de reentrada entre diferentes mapas. Ésta, en trazos muy gruesos, sería la hipótesis: la actividad de un grupo neuronal puede contribuir a la experiencia consciente si forma parte de una agrupación funcional caracterizada por la presencia de fuertes interacciones mutuas durante períodos de centenares de milisegundos. Tal agrupación tendría una composición constantemente cambiante y siempre integrada, localizándose, principalmente, en el sistema talamocortical. Según la

hipótesis del núcleo dinámico, elevados valores de una medida denominada *complejidad neuronal*, que describiría el grado en que la dinámica de un sistema neuronal es al tiempo integrada y diferenciada, son los esperables en toda agrupación neuronal que pueda ser concebida como substrato de nuestra experiencia consciente. La anatomía del sistema talamocortical, con sus masivas proyecciones paralelas de reentrada, convierten a este sistema en una base ideal para la obtención de valores elevados de complejidad neuronal (Tononi, 2004; 2008; 2012a; 2012b). La hipótesis del núcleo dinámico es, pues, la propuesta de *A Universe of Consciousness* (Edelman & Tononi, 2000) para la explicación de los rasgos definitorios de la experiencia consciente –según los autores, unicidad, privacidad, focalidad, informatividad, coherencia, dinámica y procesualidad–, pero, como señalan los autores al final del libro, una explicación científica no equivale a una experiencia subjetiva. Esto tendría que ver con la disolución o relativización de las claras delimitaciones entre problemas duros y fáciles (Chalmers, 1995) en el contexto de programas de investigación de base naturalista (García García, 2001: 292). Edelman aporta, por otra parte, una serie de argumentos destinados a apoyar su tesis de que los procesos neuronales que sustentan la experiencia consciente deben ser altamente integrados y diferenciados, es decir, que debe darse un patrón de actividad fuertemente interconexionado en áreas distribuidas del cerebro, y que ese patrón de actividad no puede ser caótico o demasiado generalizado. Así, apunta que la evidencia empírica ha venido mostrando consistentemente que la experiencia consciente se asociada con una actividad neuronal coordinada y ampliamente distribuida que vincula grupos conexos de neuronas ubicadas en distintas regiones del cerebro y que, por consiguiente, la conciencia no es prerrogativa de ninguna área particular del cerebro, sino que sus sustratos neuronales se hallan, antes bien, profusamente repartidos por el sistema talamocortical y regiones asociadas.[27] Plantea además que para sostener la experiencia consciente se requiere que un elevado número de neuronas interaccionen rápida y recíprocamente por medio del proceso de reentrada y que si estas interacciones de reentrada se bloquean desaparecen sectores enteros de la conciencia, pudiendo la propia conciencia difuminarse, dividirse o desvanecerse. Por último, se hace necesario que las pautas de actividad de los grupos neuronales que sostienen la experiencia consciente cambien constante-

[27] En este sentido, Giulio Tononi ha afirmado, en un fabuloso texto –plagado de guiños y referencias a la historia de las ciencias y las artes– que inaugura el género de la divulgación novelada en neurociencias: "The brain is a democracy –there is no such thing, in the brain, as a prince or pope, who sees and hears everything, and takes all decisions– no privileged seat of consciousness, no pontifical seat. Consciousness needs the cooperation of many specialists, each one providing its unique contribution" (Tononi, 2012a: 30).

mente y se mantengan suficientemente diferenciadas entre sí, dado que si un gran número de neuronas comienzan a dispararse sincrónicamente del mismo modo, reduciéndose la variedad y diversidad de los repertorios neuronales –tal y como sucede durante el sueño profundo o en ataques epilépticos–, la conciencia se desvanece. A las referencias al sueño y la epilepsia como sustentos contrafácticos de su teoría, Edelman suma evidencias según las cuales para que un estímulo sea percibido conscientemente es necesario que áreas del cerebro ampliamente distribuidas pero integradas presenten respuestas coordinadas y rápidas interacciones mutuas de reentrada.

Giulio Tononi, uno de los colaboradores de Edelman que participara en la elaboración de la hipótesis del núcleo dinámico, ha ofrecido recientemente una nueva hipótesis similar a aquélla en la que desarrolla un planteamiento acerca de la relación entre complejidad y conciencia que ideara con Edelman a finales de la década de los noventa (Tononi & Edelman, 1998). Según la propuesta de Tononi, que denomina *teoría de la integración de la información* y que resulta difícil decidir si habría de ser clasificada entre las teorías cognitivas o entre las neurobiológicas, la conciencia no sería sino la capacidad de un sistema para integrar información. Un sistema puede integrar información y ser así concebido dentro de este marco teórico como susceptible de portar estados conscientes en la medida en que disponga de un amplio repertorio de estados y en la medida en que cada estado de cada elemento del sistema en cuestión sea causalmente dependiente de los estados de otros elementos. La teoría de la integración de información es muy similar a la hipótesis del núcleo dinámico, en tanto propone al sistema talamocortical como base neurofisiológica de la experiencia consciente y en tanto se basa en una descripción de la dinámica neurofisiológica subyacente a la experiencia consciente en términos de su discriminatividad informativa, esto es, en términos de su diferenciación e integración. No obstante, existen diferencias entre ambas propuestas, dado que el aparato matemático utilizado para la descripción de la integración de la dinámica del sistema neuronal subyacente a la experiencia consciente es en cada caso diferente. Tononi introduce una nueva medida, Φ, la cual define como la cantidad de información causalmente efectiva que puede ser integrada a través de la más débil entre las vías de conexión de un sistema, de forma que Φ mediría interacciones causales dentro de un sistema, y cuando la dinámica de un sistema posea elevados valores de esta unidad de medida habrá de ser el mismo tenido por consciente. Tanto la *complejidad neuronal* como Φ son, en cualquier caso, medidas del equilibrio entre diferenciación e integración en la dinámica de un sistema neuronal.

*La teoría de Antonio R. Damasio*

Dado que, tal y como argumentaremos en la parte segunda, una cabal biología de la conciencia habrá de asentarse sobre planteamientos del tipo de los de Damasio, nos detendremos en una exposición de su teoría más pormenorizada que las precedentes. La misma es, de entre las elaboradas hasta la fecha, la que de forma más decidida y definida ha tratado de ubicar a la conciencia en su historia: la evolutiva. Ésta, la historia evolutiva, comienza desde luego antes de la aparición de la experiencia consciente, y Damasio trata de escrutar la lógica de dicha aparición recorriendo las sendas abiertas hacia la misma antes de que tuviera lugar. Su propósito es, pues, el de integrar la historia filogenética de la conciencia en su contexto: el de la historia de la vida. Para lograrlo atiende a los rasgos mínimos que debieron hallarse presentes en los organismos simples de los que procedemos los animales conscientes. En este sentido, plantea que en tales organismos, como las bacterias, encontramos ya la maquinaria necesaria para detectar cambios significativos tanto en su medio interno como externo, además de criterios para reaccionar de la forma oportuna ante tales cambios. Estos organismos disponen de un rasgo biológico ubicuo: un diseño homeostático, es decir, una colección de directrices destinadas a proveer los medios para mantener los parámetros vitales del organismo dentro de un rango de valores que posibilite la interina prolongación del desequilibrio termodinámico, esto es, de la vida. El mantenimiento de dichos parámetros dentro de determinados rangos de valores es la meta de todo ser vivo, y esta meta es la razón de la existencia de los señalados criterios de reacción. Pero, según Damasio (2010: cap. 2), además de estas metas y estos criterios se hacen necesarios otros ingredientes: hemos de contar con un sistema de incentivos sobre la base del cual el organismo priorice y ejecute acciones de acuerdo con sus metas homeostáticas y los criterios de reacción que de las mismas dimanan. Los incentivos y la intención, la mera intención de vivir, existían antes de que se presentaran en las mentes conscientes en forma de dolor, placer, deliberación o planes, de ahí que quepa decir que la mente consciente revela de forma paulatina los mecanismos evolutivos de la regulación de la vida que existieron largo tiempo antes de su irrupción. En los animales conscientes, los mapas neuronales de partes del cuerpo en que los parámetros de los tejidos se desvían del intervalo homeostático ha-

ciendo peligrar la integridad del organismo son experimentados como dolor o castigo, mientras los correspondientes a aquéllas cuyos parámetros se encuentran en la parte óptima del intervalo lo son como placer o recompensa. Éste es el primer paso: el mapeado neuronal del cuerpo. Pero más allá de esto, la evolución del sistema nervioso logró disponer los medios para que los organismos pudieran no sólo detectar la amenaza o la oportunidad, sino también para predecirlas. Así, por ejemplo, los eventos seguidos de recompensa se asocian a la misma desencadenando las cascadas fisiológicas de la recompensa. Lo mismo, *mutatis mutandis*, sucede con los seguidos de castigo, y en ambos casos, el objeto de las cascadas fisiológicas del castigo y la recompensa es el de optimizar la conducta, el de poner al organismo al tanto acerca de qué conviene hacer y qué conviene evitar. Esta capacidad predictiva adopta diferentes formas, pudiendo las cascadas fisiológicas de la recompensa o el castigo señalar los propios aciertos o errores predictivos, e incluso asociar estímulos o eventos entre sí y con su forma y orden de ocurrencia. En todo caso, la homeostasis y los sistemas de incentivación y predicción que en ella arraigan tienen un único propósito, el de preservar la vida, y se trata de un propósito que atraviesa la historia filogenética al completo apareciendo idénticamente en las formas de vida más sencillas y las más complejas, siendo así que, al igual que las de una arqueobacteria cualquiera, "cada operación de nuestro cerebro gira alrededor del problema de mantener la vida" (Damasio, 2004: 165). En el curso de la evolución esos sistemas de incentivación y predicción fueron haciéndose progresivamente más complejos y derivaron en lo que hoy denominamos emociones y motivaciones, y para Damasio es precisamente la emoción el substrato de toda experiencia consciente, un substrato que descansa en el modo en que nuestros sentimientos emocionales, nuestra forma de sentir las emociones, constituyen una interpretación consciente de los estados corporales. Se hace en este punto necesaria una sucinta acotación terminológica. Damasio entiende a la emoción como descendiente del valor biológico, como un refinado producto de los mecanismos ya existentes para la regulación de la vida, para la detección, evitación o aprovechamiento de riesgos o recursos. Denomina emociones a programas de acciones en buena medida automáticos, innatos, estables y predecibles que se disparan ante situaciones –en principio– biológicamente relevantes. Dichos programas incluyen cambios en el curso de la conducta manifiesta y, asimismo, en diferentes niveles somáticos, desde la actividad visceral (de intestinos, piel o corazón) a la endocrina. Por su parte, los sentimientos emocionales constituyen en su propuesta percepciones conscientes mixtas del concierto de la dinámica de las emociones, de lo que sucede en nuestros cuerpos y

nuestros cerebros cuando se desencadena una emoción. Los sentimientos emocionales pueden surgir también, aunque atenuados, de simulaciones endógenas de estados emocionales: las partes del cerebro encargadas de mapear el estado somático del organismo pueden comportarse *como si* un patrón somático típico de una determinada emoción estuviera teniendo lugar, desencadenando su correspondiente sentimiento emocional. Estos simulacros suponen un considerable ahorro de tiempo y energía por cuanto evitan la necesidad de evocar el patrón emocional completo en situaciones en las que el mismo no se hace necesario ni, a menudo, conveniente. Tanto las emociones como los sentimientos emocionales son, en último término, a la vez siervos y retoños del control homeostático, referente último de todo fenómeno biológico. Damasio colige de esta centralidad de la homeostasis que la ventaja adaptativa de la conciencia tuvo que proceder de la mejora de la regulación de la vida en entornos cada vez más complejos. Los organismos simples que regulan de forma rígida y estereotipada su conducta se encuentran adaptados a entornos poco variables y provistos de abundantes nutrientes. Dicha regulación, en el caso de requerir de intervención nerviosa, no precisa de la elaboración de mapas: basta con disposiciones, reglas del tipo "si me tocan en este costado me desplazo hacia aquél". Sin embargo, colonizar nuevos nichos resulta obviamente provechoso por cuanto amplía el acceso a recursos, pero una tal colonización sólo puede ser exitosa cuando el organismo disponga de medios para enfrentar las eventualidades propias de condiciones no necesariamente sujetas al régimen acostumbrado en su nicho de origen, esto es, cuando cuente con mecanismos en virtud de los cuales le sea dable realizar predicciones sofisticadas y ajustadas de las potenciales amenazas y recompensas y, asimismo, disponer los medios apropiados para, respectivamente, eludirlas y alcanzarlas, cuando cuente con mecanismos capaces, en resumidas cuantas, de guiar su conducta de forma flexible, y la diferencia crucial entre la regulación de la vida antes y después de la aparición de la conciencia reside en la creciente compañía que la deliberación orientada por un sí mismo unificador ha venido ofreciendo a los inflexibles automatismos, siempre conservados en el decurso evolutivo (Damasio, 2010: 176 del original, 270 de la traducción). De este modo, esos mecanismos son, en la teoría de Damasio, los pilares de la evolución de la mente y la conciencia.

Esta evolución se yergue sobre la base de los ya referidos mapas neuronales, pautas de actividad neurofisiológica que acotan cartográficamente las aferencias continua y masivamente enviadas desde todo el cuerpo al sistema nervioso central durante la interacción del organismo con su medio y que constituyen los ladrillos de la mente. Estos

ladrillos se presentan en forma de imágenes sentidas (visuales, auditivas, gustativas, táctiles, olfativas, propioceptivas, nociceptivas, termoceptivas) cuando la mente alcanza a ser consciente, cuando a la mente se le añade un sí mismo. De este modo, primero viene la mente, resultado del incesante trabajo cartográfico del encéfalo, y luego la conciencia, pues los mapas neuronales pueden perfectamente participar en la gestión de la conducta sin aparecérsele como imágenes a una mente consciente. Dichos mapas se elaboran constantemente en la interacción con el medio, pero no son una copia o una transferencia pasiva de aspectos del mismo o del estado del cuerpo, sino que diversos elementos del curso de la actividad fisiológica intervienen de forma activa en el ensamblado de los mismos. Tampoco se trata de mapas estáticos, sino en una constante fluctuación relacionada con los cambios que a cada fracción de segundo tienen lugar tanto dentro como fuera del organismo. La noción de mapa no es, por otra parte, una mera analogía: la actividad cartográfica del sistema nervioso es una de sus más conspicuas propiedades. La forma más común de ilustrarla es la retinotopía, el modo en que la actividad neurofisiológica en la corteza occipital conserva las relaciones geométricas habidas entre los fotones que inciden en diversos puntos de la retina. Aunque este ejemplo sea el más claro, la cartografía dinámica, constructiva e interactiva del propio cuerpo y el entorno es la tarea que mantiene entretenida a buena parte de nuestros sistemas nerviosos centrales. Este trajín cartográfico no se detiene en el mapeado del propio cuerpo y el entorno, sino que determinados segmentos de ciertos sistemas nerviosos centrales complejos construyen mapas de sí mimos elaborando mapas, llegando las imágenes que los mismos soportan a adquirir un elevado grado de abstracción. En cualquier caso, diferentes partes del sistema nervioso central confeccionan distintas clases de mapas. Así, por ejemplo, la corteza cerebral produce mapas detallados, mientras determinadas estructuras subcorticales, como los colículos, el núcleo del tracto solitario y el núcleo parabraquial, trabajan con mapas más toscos. Algunas áreas del encéfalo producen de forma directa los mapas que conforman el substrato de lo mental, otras contribuyen apoyándolas y otras, sencillamente, no ocupan lugar alguno en la tarea de dar origen a la vida mental. La médula espinal cae dentro de esta tercera categoría, del mismo modo que lo hacen el cerebelo y el hipocampo, que no obstante contribuyen al control del movimiento, la memoria y la emoción, pero cuya abolición no supone la de la dinámica de imágenes en que consiste el proceso que denominamos “mente”. Las dos estructuras troncoencefálicas aludidas, el núcleo del tracto solitario y el núcleo parabraquial –y, con ellas, la sustancia gris periacueductal, tan profusamente conectada con dichas estructu-

ras como las mismas lo están a su vez entre sí–, son, por su parte, de capital importancia en la propuesta de Damasio, dado que ellas reciben señales que describen el estado del medio interno del cuerpo al completo y elaboran mapas constantemente actualizados acerca del mismo. La actividad de estos mapas, producidos en el incesante bucle de la estrechísima interacción entre el cuerpo y las referidas estructuras troncoencefálicas, es la que da origen al primer destello evolutivo de la conciencia, las sensaciones o sentimientos primordiales. El portugués aporta gran cantidad de evidencia neurofisiológica y neuropsicológica en defensa de esta tesis. Destaquemos de pasada la relativa a la presencia de sensaciones elementales de dolor y placer en pacientes cuyas cortezas insulares han sido destruidas, restando sólo la posibilidad de que tales sensaciones se produzcan en los apuntados núcleos troncoencefálicos –tal y como Damasio (2010) arguye en la octava nota al pie del capítulo tercero–. Damasio (Ibíd.: 167 del original, 256-257 de la traducción) justifica la centralidad de las emociones en su teoría apuntando que las mismas ofrecen las más certeras entre las indicaciones acerca de la naturaleza de la conciencia por cuanto su ejecución real corre a cargo de la sustancia gris periacueductal, íntimamente ligada al núcleo del tracto solitario y el núcleo parabraquial, las estructuras encargadas de la generación de los mapas somáticos de cuya dinámica derivan los primordiales sentimientos emocionales. Además, las emociones juegan en la mente un papel similar al de la presión selectiva en la teoría de la evolución: de acuerdo con la hipótesis del marcador somático (Damasio, 1996), ellas introducen un elemento valorativo que resulta esencial para que unos mapas neuronales preponderen sobre otros, para establecer grados de saliencia relativos a la relevancia del mapa en cuestión para la gestión de la vida y, así, grados de influencia correlativos en el proceso de conformación de una mente. El sistema nervioso central genera, por otra parte, diferentes tipos de mapas: los interoceptivos, referidos al estado del interior del organismo (musculatura lisa) y a los parámetros de estado del medio interno; los propioceptivos, referidos al estado de otros aspectos del propio cuerpo (estado de articulaciones, musculatura estriada y determinadas vísceras); y los exteroceptivos, referidos a los objetos externos que afectan al cuerpo. Los primeros son básicos e imprescindibles: en ellos estriba la primordial sensación de ser, y las sensaciones son en Damasio una clase fundamental de imágenes en cuya esencial relación con el cuerpo reside el motivo por el cual cualquier otra clase de imagen llega a ser sentida, pues ello se debe a que, cuando estas otras clases de imágenes son, efectivamente, sentidas, no lo son sino en virtud de su ir acompañadas de sensaciones. Lo apuntado en este párrafo debe leerse a la luz del anterior, porque esta

dinámica neurocartográfica sólo cobra sentido desde la perspectiva de la gestión de la vida, a la que contribuye mediante la detección y predicción de riesgos y oportunidades. Damasio sostiene que cuando el organismo puede valerse del producto de esta dinámica neurocartográfica, la refinación y el alcance de esa gestión se ven ampliados, y más cuando al mismo se superpone la experiencia sentida en el contexto de la conducta compleja (Damasio & Carvalho, 2013: 145), y más aún cuando cabe la posibilidad de integrar de forma sutil mapas de diferentes modalidades, e incluso más cuando cabe la de llevar a término dicha integración con los mapas almacenados que denominamos registros mnemónicos, posibilidad que abre las puertas de la ideación imaginativa y la planificación deliberativa.

Antes de presentar por vez primera al gran público su teoría de la conciencia, en el libro de 1999 *The Feeling of what Happens*, Damasio había distinguido ya dos formas de conciencia: la nuclear y la extendida. Un año antes de la publicación de dicho libro, Damasio explicaba en los siguientes términos el sentido de su distinción:

> Core consciousness corresponds to the transient process that is incessantly generated relative to any object with which an organism interacts, and during which a transient core self and transient sense of knowing are automatically generated. Core consciousness requires neither language nor working memory, and needs only a brief short-term memory. Extended consciousness is a more complex process. It depends on the gradual build-up of an autobiographical self, a set of conceptual memories pertaining to both past and anticipated experiences of an individual, and it requires conventional memory. Extended consciousness is enhanced by language." (Damasio, 1998: 1879).

La distinción se asemeja a la de Edelman entre una forma primaria y una secundaria de conciencia y guarda asimismo relación con la habitual asunción en la neurobiología de la conciencia según la cual la conciencia fenoménica es anterior y más básica que la conciencia de acceso. En esta línea, Damasio defiende que la conciencia extendida depende de la nuclear, anterior y más básica, y ésta, a su vez, de la integridad del *proto sí mismo* (proto-self) sujeto de los primeros destellos de la experiencia consciente, los sentimientos primordiales de origen troncoencefálico.

La conciencia nuclear es en el planteamiento de Damasio una forma de actividad mental que permite al organismo sentir como suyos los contenidos de su mente, una forma de actividad mental dotada de la perspectiva de la primera persona. Esta perspectiva, en el caso de la conciencia nuclear, se restringe a la experiencia de estar presente en el inmediato aquí y ahora, sin concomitancias conceptuales de tipo lingüístico, pues la conciencia nuclear se presenta, de hecho, como previa al lenguaje y, en cierto sentido,

como una condición de su posibilidad: Damasio afirma que el lenguaje probablemente no hubiera evolucionado en ausencia de organismos dotados de esta forma de conciencia y se pregunta cuál podría ser la utilidad del mismo para criaturas desprovistas de la misma, marcando así distancias con teorías como la de Gazzaniga, aplicables exclusivamente, en su opinión, a formas de conciencia de tipo autobiográfico como la que describiremos en el siguiente párrafo. La conciencia nuclear, en la última formulación de Damasio, resulta de la interacción del organismo tal y como aparece mapeado en las estructuras troncoencefálicas en las que arraigan los sentimientos primordiales con otras partes del encéfalo encargadas de mapear la dinámica de la relación entre el organismo y objetos externos. Partimos pues, del *proto sí mismo* y sus sentimientos primordiales, la forma más básica de la mente sentida, una descripción neuronal del estado somático del organismo. A éste se añade el sí mimo central (el yo de la conciencia nuclear) cuando las modificaciones del organismo en interacción con objetos externos dan lugar a transitorios entrelazamientos funcionales de mapas referidos propiamente a los objeto externos y mapas del *proto sí mismo*. El esqueleto del sí mismo central lo integran de este modo imágenes del estado del organismo y de las modificaciones sufridas por el *proto sí mismo* en interacción con los objetos externos, imágenes de las respuestas emocionales producidas en esa interacción e imágenes de dichos objetos: se trata de un esqueleto que tiene la forma de la conexión entre el organismo y los objetos externos. Esta nueva forma del sí mismo incluye la perspectiva en que los objetos son acotados en mapas, el sentimiento de que los objetos están siendo representados en una mente que pertenece al organismo, el de agencia y, finalmente, los insoslayables sentimientos primordiales, incesante testimonio de la cruda presencia de un cuerpo vivo, un testimonio que oscila entre el primer plano y la más queda de las presencias y se halla extendido entre los dominios del placer y el dolor. La esencia de la experiencia consciente es para Damasio la sensación de ser, la sensación del yo. La conciencia nuclear se alza sobre el *proto sí mismo*, la sensación de ser más básica. La modificación de los mapas del *proto sí mismo* como resultado de su integración con mapas relativos a eventos acaecidos en la relación del organismo con su entorno daría lugar a la emergencia de un mapa de segundo orden referido al modo en que el *proto sí mismo* se ve afectado por tal relación. Este mapa de segundo orden constituye una representación integrada del *proto sí mismo* con los mapas relativos a objetos externos y las relaciones causales entre éstos y aquél, la forma de representación que conforma la conciencia nuclear.

La conciencia extendida, por su parte, excede en el plano temporal el alcance de la conciencia nuclear, pues abarca una serie de estados mentales que no se encuentran ya restringidos al presente del organismo, sino que tienen por objeto asimismo su pasado y su futuro. La emergencia de esta clase de estados mentales se alza sobre la de un yo autobiográfico –concebido como colección individual de memorias y expectativas–. A pesar de que se trata de una nueva forma en la que los contenidos mentales se hacen presentes para un individuo, Damasio subraya que la conciencia extendida desempeña a su nivel exactamente el mismo papel que la conciencia nuclear al suyo: el de conferir a los contenidos mentales la perspectiva del individuo, el de hacérselos suyos. Para hacer patente la relación de dependencia entre ambas formas de conciencia, Damasio señala que el deterioro de la conciencia extendida no afecta a la conciencia nuclear, como puede entenderse que sucede en el caso de pacientes con pérdidas severas de memoria autobiográfica, mientras que lo contrario ocurre cuando es la conciencia nuclear la afectada: durante una crisis epiléptica, un estado de coma o el sueño profundo no hay ninguna forma de conciencia y ningún estado mental se presenta como suyo al sujeto. La cadena de dependencia se extiende, por supuesto, al *proto sí mismo*, en cuya ausencia ninguna de las formas posteriores de conciencia pueden tener lugar. Para que surja el sí mismo autobiográfico (el yo de la conciencia extendida) hemos de contar con los mecanismos talamocorticales en que se asienta el sí mismo central: deben, por una parte, hallarse disponibles recuerdos biográficos tratados como objetos en pulsos del sí mismo central cuyas imágenes, por otra parte, puedan interactuar ordenadamente con los mecanismos del *proto sí mismo* y, finalmente, los resultados de esta interacción han de mantener cierta coherencia dentro de una determinada ventana temporal. Esta interacción y esta coherencia dependen de mecanismos de coordinación en virtud de los cuales las imágenes mnésicas puedan organizarse actualmente y ser entregadas a los resortes del *proto sí mismo*, unos mecanismos que intervendrán también en el almacenado de los resultados de esta orquestación. Damasio apunta a los núcleos asociativos del tálamo como uno de estos mecanismos de coordinación, y asimismo a las regiones de convergencia y divergencia ubicadas en las cortezas de asociación –principalmente en las cortezas polar y medial del lóbulo parietal, las mediales prefrontales, las confluencias temporoparietales y las cortezas posteromediales– y el claustro,[28] ricamente conectado con las aludidas

[28] No podemos pasar sin mencionar que esta estructura atrajo la atención de Francis Crick durante los últimos compases de su carrera. Su rica conectividad le hizo convencerse de su centralidad para la explicación del fenómeno de la experiencia consciente. Sólo tres semanas antes de morir le dijo a Ramachan-

regiones de convergencia y divergencia, particularmente con la corteza posteromedial, la cual describe como una región de convergencia y divergencia de alto nivel por cuanto recibe *inputs* de las cortezas de asociación parietales y temporales, cortezas entorrinales, frontales, cinguladas, la amígdala, regiones premotoras y núcleos talámicos, intralaminares y dorsales (esto es, de la práctica totalidad del encéfalo, exclusión hecha de las cortezas sensoriales y motoras primarias) y envía *outputs* hacia prácticamente todos los lugares en que convergen los *inputs* proyectados hacia ella.

De lo señalado puede colegirse que la teoría de Damasio no pretende reducir la conciencia a una forma superior de cognición, como sucede con las teorías representacionales de pensamiento de orden superior y las teorías cognitivas, sino que plantea la existencia de formas sencillas de experiencia consciente que habrían surgido pronto en la historia filogenética y que no requieren de operaciones cognitivas complejas como las propias de la atención, la memoria de trabajo, el razonamiento o el lenguaje (Damasio, 1999: 107-125; 184-189). La conciencia extendida tendría su origen, precisamente, en la confluencia entre un funcionamiento normal de estas capacidades superiores y el sentimiento del aquí y ahora en que consiste la conciencia nuclear.

En su último libro hasta la fecha, *Self Comes to Mind: Constructing the Conscious Brain*, Damasio, a pesar de que comienza por advertir que pretende en el mismo reelaborar su concepción de la biología de la conciencia, amplía su tratamiento del tema sin modificaciones sustantivas del marco teórico en que había venido trabajando, aunque, ciertamente, en la década que separa *The Feeling of what Happens* de *Self Comes to Mind* el protagonismo concedido por Damasio al sí mismo primordial, el *proto sí mismo*, no hizo sino aumentar hasta aparecer el mismo como la sensación fundamental e independiente en su origen –a priori– de la relación del organismo con cualquier clase de objeto, como el elemento decisivo del sí mismo, subyacente a cualquier forma de experiencia.[29] Sea como fuere, el marco teórico general, como indicábamos, se ha mantenido, emplazándose idénticamente a su base la idea de que la conciencia hace su aparición en la historia evolutiva cuando a los procesos mentales básicos se les añade el proceso del

dran que sospechaba que en ella reside el secreto de la conciencia (Ramachandran, 2004b: 1154). Sin duda, la lectura de Crick de los recientes resultados de un estudio de caso único de estimulación eléctrica del claustro izquierdo y la ínsula anterodorsal (Koubeissi et al., 2014) habría sido tan excitante para él como para el resto de los investigadores en los *Consciousness Studies*. Como su más estrecho colaborador ha explicado (Koch, 2014: 26-27), lo excepcional del estudio es que por primera vez se encuentra una zona concreta (la sustancia blanca contigua al claustro) cuya estimulación eléctrica puede abolir abruptamente la conciencia, reestableciéndose la misma con el cese de la estimulación.

[29] En vista de este aumentado protagonismo, la crítica a la teoría de Damasio en el último capítulo de Denton (2005) pierde todo el fuelle que hubiera cabido presumirle.

yo o sí mismo. El primer peldaño de la escala evolutiva de la conciencia sigue asimismo ubicándose en el *proto sí mismo*, que sigue definiendo el portugués como un proceso acaecido en virtud de un tipo especial de imágenes mentales del cuerpo del organismo producidas en estructuras troncoencefálicas encargadas de mapear el estado del cuerpo. Frente a la unilateralidad talamocortical imperante en la neurobiología de la conciencia, esta insistencia en el papel fundamental del tronco del encéfalo en la génesis de la experiencia consciente no pretende desestimar la a día de hoy incuestionable contribución del tálamo y la corteza en el ensamblaje de la experiencia consciente, sino sólo llamar la atención sobre la extremadamente probable prioridad funcional de diversos mecanismos troncoencefálicos en dicho proceso. La relación entre estas primitivas estructuras neuronales –dedicadas a la cartografía somática y ubicadas en las regiones superiores del tronco del encefálo– y el cuerpo del organismo aparece en esta ocasión enfatizada, al punto que la posibilidad de trazar fronteras entre aquéllas y éste se presenta problemática. Damasio incide en este sentido en que la idea más importante entre las contenidas en *Self Comes to Mind* es la de que el cuerpo es el fundamento de la experiencia consciente. Asentado de este modo en el cuerpo, el *proto sí mismo* constituye la base sobre la que se alzan las sucesivas formas de conciencia, y sus productos más elementales son los sentimientos primordiales, que reflejan el estado del cuerpo proporcionando una experiencia directa del mismo presente de manera continua y espontánea durante todo el tiempo que un organismo es consciente. Es necesario destacar que Damasio insiste en que estos sentimientos primordiales difieren en su fenomenología y su neurofisiología de las experiencias perceptivas, de forma que, desde su punto de vista, el basamento de la experiencia se encuentra antes en los estados emocionales que resultan de las modificaciones somáticas y las tienen por objeto que de la relación perceptiva entre el organismo y su medio, una relación cuyo ser sentida arraiga igualmente en estos sentimientos primordiales. Es precisamente en la relación con los objetos del entorno donde se origina, como veíamos, el siguiente nivel de la escalada evolutiva de la conciencia, pues la conciencia nuclear nos es presentada como vinculada con la acción del organismo en su medio en tanto la misma se genera cuando el *proto sí mismo* se ve modificado por la interacción entre el organismo y el objeto. A esta tipología de formas de conciencia viene a sumarse, finalmente, el yo autobiográfico o extendido. Cada uno de los tres elementos de esta tipología, el *proto sí mismo*, la conciencia nuclear y la extendida o autobiográfica, se asientan en diferentes formas de actividad neurobiológica, aunque todas ellas cortadas por el patrón de la conectividad reentrante masiva y la activación sincró-

nica ubicuas en la neurobiología de la conciencia. El elemento común a las tres formas de conciencia de las que habla Damasio es el sí mismo, la sensación de mismidad, sucesivamente perfilada en cada uno de los tres niveles. Pero no debemos entender que Damasio pretenda describir *cosas* discretas mediante categorías rígidas: las demarcaciones entre las diversas formas del sí mismo pueden concebirse como fluidas hasta un extremo en que cabría definirlas como poco menos que artefactos retóricos.

### 4._Coda

Exceptuando el caso de Changeux (2002; 2008), que critica severamente hipótesis como las de Crick o Edelman por considerarlas simplistas e insuficientemente justificadas, tanto desde el punto de vista empírico como desde el teórico, y, particularmente, por no "proponer ninguna implementación informática de sus hipótesis en términos de redes de neuronas formales" (Changeux, 2008: 224 del original, 179 de la traducción), aunque tampoco aclare los motivos por los cuales tal cosa habría de ser necesaria, son escasas las referencias cruzadas que en la neurobiología de la conciencia hacen pensar en un contexto en el cual se busque el imperio y son, asimismo, pocos los que entienden que su teoría será capaz de explicar, en solitario, la conciencia al completo, en toda su extensión, todos sus rasgos y aspectos y todas sus presumibles formas, dispersas por todo lo ancho y largo del reino animalia. La presencia de sucesivos niveles explicativos, la necesidad de complementariedad y la inevitabilidad de solapamientos teóricos se les escapan hoy, pues, a pocos. Incluso teóricos tan desemejantes como Damasio y Gazzaniga comentan mutua y elogiosamente sus respectivas propuestas sin pretender la discrepante incompatibilidad de las mismas (vid. Gazzaniga, 2008: 279 del original, 289-290 de la traducción; 320 del original, 329 de la traducción; Damasio, 2010: 208 del original, 311 de la traducción). Parecen comprender que la conciencia es un fenómeno con muchas capas e igualmente muchas habrán de ser las de su explicación. Algunas de las teorías neurobiológicas que hemos comentado tienen por objeto procesos cognitivos de alto nivel y habremos de tenerlas presentes a la hora de explicar el modo en que la conciencia se relaciona con los mismos; otras, más cercanas a procesos afectivos, serán de utilidad para integrar en los *Consciousness Studies* la ineludible guía de la psicobiología comparada de cara a abordar la filogénesis de la experiencia consciente. Habremos de precisar estas ideas en la segunda parte. Por el momento, basta con ponerlas sobre la mesa y pasar ya a encargarnos de bloquear el desafío misteriano, un desafío según el

cual cualquier clase de esfuerzo teórico del tipo de los reseñados en este capítulo será por siempre vano. En el próximo capítulo sometemos a crítica los argumentos misterianos. En la segunda parte discutiremos acerca de los medios más propicios para el avance hacia aquello que los misteiranos niegan que pueda alcanzarse: una explicación científica de la conciencia.

# CAPÍTULO 6

## ARGUMENTOS MISTERIANOS. MOTIVOS PARA DECIR NO AL NO APRIORÍSTICO

> *Consciousness is a fascinating but elusive phenomenon; it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written about it* (Sutherland, 1989).

La citadísima frase de Stuart Sutherland con la que abrimos este capítulo puede sonar un tanto informal para un prestigioso diccionario de psicología, pues precisamente en un diccionario de psicología fue publicada, pero, en cualquier caso, nos es útil aquí para enmarcar nuestra discusión de los argumentos esgrimidos por unos cuantos filósofos que, podría decirse, han venido tomándose verdaderamente en serio, e incluso ampliando, el sentido de la palabra «imposible» en el contexto de esta cita y, ante todo, tratando de demostrar que precisamente ésa es la palabra a utilizar en semejante contexto. Estos filósofos no se muestran en absoluto conformes cuando neurocientíficos de primera línea se refieren a la conciencia utilizando palabras tan enfáticas y altísonas como, por ejemplo, "el mayor de los problemas irresueltos en toda la biología" (Albright et al., 2000: 40τ). «Problema» les sabe a poco, e «irresuelto» también, pues se proponen persuadir a la comunidad científica del carácter *irresoluble* del problema de la conciencia y, por ende, del sentido en que debiéramos dejar de hablar del problema de la conciencia y sustituir esa tímida locución por "el *misterio* de la conciencia" –un giro, sin duda, mucho más efectista.

Cuando Max Planck ingresó en la Universidad de Munich en 1875, su profesor de física le aconsejó que se dedicara a otra materia, pues ya no había nada que descubrir en

física. Veinte años más tarde Robert Millikan recibió el mismo consejo de manos de otro futurólogo incompetente. Los misterianos abundan aún hoy en la futurología pesimista del profesor de Planck. A la exposición y crítica del modo en que vienen haciéndolo dedicaremos los siguientes apartados.

## 1._Introducción

> *And then there is the brand of defeatism in my own home discipline, philosophy of mind, the* mysterian *doctrine that insists that the human brain is simply not up to the task of understanding the human brain, that consciousness is not a puzzle but an insoluble mystery (so stop trying to explain it). What is transparent in all these claims is that they are not so much defeatist as protectionist: don't even try, because we're afraid you might succeed!* (Dennett, 2006: 260 del original, 305 de la traducción.)

Durante décadas, desde la etapa fundacional entre los cuarenta y los cincuenta hasta bien entrados los ochenta,[1] la conciencia fue concebida dentro de las distintas áreas de las ciencias cognitivas como algo tan esotérico y extraño, algo cuyo posible tratamiento científico resultaba tan remoto y hasta inimaginable que la tendencia consistió en mirar para otra parte cuando surgían las perplejidades y dificultades teóricas, conceptuales y metodológicas a las que el estudio de los fenómenos relacionados con la conciencia aboca y, asimismo, y tal y como en los albores de los hoy denominados *Consciousness Studies* señalara Dennett (1978), en dejar que fueran los filósofos quienes se embarcaran en solitario en la empresa de hacer el ridículo tratando de capturar a la cierva de Cerinea de la conciencia e integrarla en una teoría respetable.[2] La actitud de la abrumadora mayoría de los investigadores en las diversas áreas de las ciencias cognitivas hacia las posibilidades del estudio científico de la conciencia ha experimentado en las casi cuatro décadas transcurridas desde el sutil diagnóstico de Dennett un giro desde la desconfianza, el recelo y las reservas hacia un optimismo investigador evidenciado en

[1] Década a lo largo de la cual aparecieran importantes trabajos científicos sobre la conciencia, como los de Johnson-Laird (1983a; 1983b), Jackendoff (1987), Baars (1988) o Edelman (1989), de los que hablamos en el capítulo anterior, e incluso compilaciones como la de Marcel & Bisiach (1988), aparecida tres años después de un congreso organizado por los propios Tony Marcel y Edoardo Bisiach para celebrar el retorno de la conciencia a las ciencias cognitivas.

[2] Aludimos aquí al siguiente fragmento, adaptando libremente al castellano alguno de sus segmentos: "Consciousness appears to be the last bastion of occult properties, epiphenomena, immeasurable subjective states –in short, the one area of mind best left to the philosophers. Let them make fools of themselves trying to corral the quicksilver of 'phenomenology' into a respectable theory" (Dennett, 1978: 149).

un caudal de publicaciones sobre el tema que es actualmente constante, exuberante, plural y multidisciplinar. No obstante, este optimismo investigador no parece haber calado en determinados círculos filosóficos. La afortunada acuñación por parte de Owen Flanagan (1984/1991) de la noción de *nuevos misterianos* (new mysterians) responde, justamente, a la necesidad de delimitar el perímetro de esos círculos mediante una sola signatura capaz de dar elocuente cuenta de las perspectivas compartidas por los filósofos que rondan dentro de los mismos. Frente al señalado entusiasmo investigador y frente al optimismo respecto de la posibilidad de avanzar hacia una ciencia de la experiencia consciente, en estos círculos misterianos se respira la convicción de que semejante empresa científica está, por principio, abocada al fracaso. Con todo, a pesar de lo afortunado de la acuñación de Flanagan, ésta puede ser matizada y utilizada de acuerdo con criterios diversos de los propuestos con ella por el autor. Un primer paso hacia esa matización consistiría en poner de manifiesto aquello que no pretendemos matizar: el planteamiento de Flanagan se apoya en la distinción que Chomsky (1976) trazara entre problemas –cuestiones que pueden ser comprendidas y, virtualmente, resueltas– y misterios –cuestiones tan incomprensibles como irresolubles–. La distinción puede resultar difícil de justificar, pero aquí sólo estamos tratando de crear un marco léxico, y el sentido en que «misterio» se halla en la raíz de lo que aquí denominaremos «filósofos misterianos» será precisamente el sentido en el que Chomsky planteó dicha noción; esto es, hablaremos de misterianismo en el ámbito de los *Consciousness Studies* para designar tentativas destinadas a establecer el carácter misterioso, en el aludido sentido chomskiano, de la conciencia. Hasta aquí acompañamos a Flanagan. Su noción de *nuevos* misterianos, en cambio, nos es innecesaria. Expongamos brevemente por qué. Flanagan (1992: 9) distingue entre nuevos y viejos misterianos, atribuyendo a estos últimos una concepción dualista según la cual la conciencia no puede ser explicada científicamente porque no constituye un fenómeno natural, mientras los nuevos misterianos, según la distinción de Flanagan, defenderían la existencia de la conciencia como un fenómeno natural pero científicamente inabordable e inexplicable. Nuestra exposición no necesitará de esta distinción entre nuevos y viejos misterianos. Así, haremos uso sólo de la mitad de la signatura de Flanagan, hablando, pues, de misterianos en sentido lato, esto es, refiriéndonos con esta noción a aquellos teóricos que niegan la posibilidad de alcanzar una explicación científica de la conciencia, ya se trate de una negación planteada por autores con preferencias por metafísicas monistas o dualistas, naturalistas o supernaturalistas. En este sentido, nuestra opción terminológica obedece no sólo al hecho de que los pro-

pios filósofos misterianos suelan presentar sus planteamientos en términos antes epistemológicos que ontológicos (De Caro, 2009: 499), resultando así secundarias sus querencias ontológicas cuando se trata de sus inclinaciones misterianas, sino, decisivamente, a que, aun cuando no podamos entender aquéllas como secundarias por el influjo que las mismas hayan podido ejercer sobre éstas, la línea divisoria entre viejos y nuevos misterianos, entendemos, ha venido haciéndose problemática conforme el debate acerca de la conciencia se fragmentaba y hacía cada día más complejo, acogiendo cada vez más facciones enfrentadas en más y más sub-discusiones sutil y problemáticamente interconectadas. Más allá de esta línea divisoria, y para fijar en una definición concisa qué entenderemos por «misterianismo», podemos, con Uriah Kriegel, describirlo en pocas palabras como "la idea de que existen razones sistémicas y de principio que hacen humanamente inalcanzable una satisfactoria explicación de la conciencia" (Kriegel, 2009: 461τ).

El misterianismo ha venido constituyendo una postura filosófica discutida por filósofos y de interés para científicos cognitivos con inclinaciones filosóficas, aunque, ciertamente, ha logrado exceder las fronteras de la filosofía para, por ejemplo, sorprender y frustrar a Dennett (2003: 20 del original, 36 de la traducción) con las simpatías de Pinker (1997: 561 del original, 715 de la traducción) hacia esta postura, e incluso para llegar al gran público convertido en escepticismo y suspicacia respecto de las neurociencias, la psiquiatría, la psicología y las ciencias cognitivas en general de la mano de, por ejemplo, el exitoso periodista científico John Horgan (Horgan, 2000) –que previamente habría abordado una tarea aun mayor: la de contagiar de pesimismo no ya respecto del desarrollo de las neurociencias, que en la obra citada describe como un antiprogreso, sino respecto de las posibilidades del desarrollo futuro de la propia ciencia (Horgan, 1996)–. Los misterianos, como vemos, nos salen al paso procedentes de casi cualquier lugar. Aquí nos ocuparemos sólo de los misterianos procedentes de facultades de filosofía, y particularmente, claro, de sus argumentos.

Antes de seguir adelante y entrar en detalles, quisiéramos anticipar el rasgo común a las diversas variedades de argumento misteriano, y encontramos que no cabe mejor forma de hacerlo que aludiendo a la noción de *argumento hegeliano*. La misma ha sido recientemente acuñada por Anthony Chemero en los primeros compases de su libro de 2009 *Radical Embodied Cognitive Science*. Con ella se refiere Chemero (2009: 4 y ss.) a argumentos basados en escasa o nula evidencia empírica que pretenden probar a priori la inadecuación de un proyecto o incluso un programa de investigación empírica al

completo. Este tipo de argumento abunda en las disciplinas jóvenes, como las ciencias cognitivas, pero, afortunadamente, según Chemero, los mismos suelen tener tanto éxito como sustento empírico: es raro el equipo de investigación que abandona su proyecto después de leer un argumento hegeliano en un artículo publicado en el *Philosophical Quarterly*. No obstante, Chemero indica también, al final del libro al que venimos aludiendo, que, más allá de la formulación y discusión de argumentos hegelianos, hay un trabajo constructivo que los filósofos interesados en las ciencias cognitivas pueden llevar a cabo, y que este trabajo puede de hecho resultar muy útil para el desarrollo de las mismas. Más allá del contraproducente intento de demostrar imposibilidades, dicho trabajo incluiría, por ejemplo, mostrar, mediante el esclarecimiento de enredos conceptuales, el modo en que determinados desarrollos son de hechos posibles; plantear cómo y por qué las perspectivas dentro del área son o no coherentes; situar los problemas actuales de la disciplina en perspectiva histórica o presentar a los investigadores el modo en que su trabajo se apoya en cuestiones filosóficas (Chemero, 2009: 207). Sobra incidir en que no es éste el tipo de trabajo filosófico que comentaremos y criticaremos a lo largo de este capítulo. Sin embargo, y a pesar de que la crítica de argumentos hegelianos no se encuentre en la lista de Chemero de trabajos constructivos que los filósofos pueden realizar en ciencias cognitivas, entendemos, con Andrew Brook (Brook, 2005: 398), que esta clase de argumentos, antes que ignorados, deben ser respondidos y criticados.

En los próximos apartados expondremos los argumentos que desde los señalados círculos misterianos han venido esgrimiéndose en defensa de esa intuición según la cual todo esfuerzo invertido en la tarea de desentrañar científicamente la experiencia consciente será, en cualquier caso, vano. Tras su exposición, emprenderemos la crítica de dichos argumentos, desvelando la inconsistencia de las conclusiones a las que arriban los misterianos desde sus diversas, aunque análogas, intuiciones de partida.

## 2._Los argumentos misterianos

Trazaremos una distinción de grado antes de comentar los diferentes argumentos misterianos. Vamos a ocuparnos de una serie de argumentos destinados a demostrar la imposibilidad de la elaboración de una ciencia de la experiencia consciente.[3] No obstan-

[3] Hay que matizar que dichos argumentos son en muchas ocasiones presentados por sus proponentes como refutación de una concepción "fisicalista" de la conciencia fenoménica, esto es, una concepción

te, aun cuando todos y cada uno de los señalados argumentos comparten este objetivo y pretenden haber hallado o suponer insalvable óbice para dicha elaboración, tendremos que hablar, por así decir, de grados, dado que no todos estos argumentos intentan obliterar la vía hacia *cualquier* posible desarrollo científico capaz de ofrecer explicación de la experiencia consciente, sino que algunos pretenden sólo probar la imposibilidad de una tal explicación en caso de que fuera formulada desde un marco cognitivista o representacional. Así, podemos hablar, en primer lugar, de un grupo de argumentos misterianos absolutos o radicales, que comparten el conato de negar a priori la posibilidad de cualquier clase de explicación científica de la experiencia consciente. Dentro de este primer grupo de argumentos encontramos el argumento epistemológico (Jackson, 1982) y el del punto de vista (Nagel, 1974), así como el de la brecha explicativa (Levine, 1983) y el de la clausura cognitiva (McGinn, 1989). En segundo lugar, entendemos, debe hablarse de un conjunto de argumentos misterianos parciales o anticognitivistas, destinados, como sugeríamos, a demostrar el impasse en que se encontrarán los intentos explicativos cognitivistas o representacionales. Los argumentos misterianos de esta segunda clase se basan en una serie de supuestos acerca de las implicaciones –pretendidamente nefastas para cualquier tentativa cognitivista o representacional– que traería consigo el hecho de que podamos concebir una serie de escenarios ficticios o hipotéticos. Describiremos estos escenarios y nos adentraremos en las supuestas implicaciones de nuestra supuesta capacidad para concebirlos después de pasar revista, a renglón seguido, al aludido primer grupo de argumentos, aquéllos a los que nos hemos referido como misterianos absolutos o radicales. Dedicaremos, pues, los siguientes subapartados a presentar los argumentos misterianos y el resto del capítulo a analizarlos críticamente.[4]

---

según la cual ninguna entidad inmaterial y por principio ajena al alcance del escrutinio de las ciencias naturales es necesaria para la existencia de la experiencia consciente. Cabría sugerir que rebatir una ontología fisicalista no equivale a rebatir la posibilidad epistémica de ofrecer una explicación científica de la conciencia. No obstante, hasta que alguien acierte a aclarar el modo en que una explicación de la conciencia que necesite postular entidades ajenas al mundo físico e inmunes al escrutinio de las ciencias naturales puede ser una explicación científica, dicha equivalencia permanecerá incólume.

[4] Destaquemos que, ciertamente, nuestra opción a la hora de distinguir entre diversas clases de argumentos misterianos es una entre otras: hemos optado por ella por mor de la claridad expositiva, dado que nos interesa, ante todo, distinguir entre los argumentos destinados a bloquear la posibilidad de una explicación científica *cualquiera* de la conciencia y los argumentos destinados a bloquear la posibilidad de una determinada clase de explicación científica de la conciencia (los argumentos que incluimos en este segundo grupo se dirigen, como apuntábamos, contra el computacionalismo, el funcionalismo, el representacionalismo y el cognitivismo). Se trata, así, de una opción que obedece a intereses expositivos (*ordo artificialis*). Podríamos, por ejemplo, haber organizado nuestra exposición en torno a la habitual distinción entre un tipo epistemológico y otro ontológico de misterianismo (vid., v. g., Kriegel, 2007; 2009). Otra posibilidad interesante hubiera consistido en distinguir entre misterianos, digamos, "irrestrictos", según los cuales el problema de la conciencia es por principio y definitivamente irresoluble, y misterianos revolucionarios o "expectantes", según los cuales cabe esperar que una revolución en las ciencias perti-

## 2.1._Argumentos misterianos absolutos o radicales

Dentro de este primer bloque nos ocuparemos, en primer lugar, del argumento epistemológico y el argumento del punto de vista. Seguidamente hablaremos de la brecha explicativa y concluiremos presentando la tesis de la clausura cognitiva.

### *a) El argumento epistemológico y el argumento del punto de vista*

Tanto el argumento del punto de vista como el epistemológico constituyen verdaderos puntos de referencia en el área de los *Consciousness Studies* dado que, planteémoslo en estos términos, pusieron en circulación las metáforas adecuadas en el momento adecuado. Se trata de dos memes que han tenido un enorme éxito, dos argumentos tan relevantes y comentados que, o bien sobran las presentaciones, o bien, si pretendieran ellas hacer justicia a los mismos tratando de dar acabada cuenta de los motivos del señalado éxito, acabarían resultando excesivamente prolijas. En semejante ínterin, puede que lo más prudente sea pasar directamente a la exposición de ambos argumentos con la vista puesta en su posterior crítica. Mas, ¿por qué hemos agrupado ambos argumentos bajo un epígrafe conjunto? Baste indicar que autores como Lewis (1983c), McMullen (1985) o Pereboom (1994) entienden que ambos configuran una misma línea argumentativa, llegando algunos a hablar del "Nagel-Jackson Knowledge Argument" (vid. Bealer, 1994; Pereboom, 1994).[5]

---

nentes ofrezca la posibilidad de resolverlo. Los primeros pretenden, por así decir, saber de antemano qué podremos y qué no podremos llegar saber, mientras los segundos, más prudentes, proponen que, en el estado actual de las disciplinas científicas pertinentes, el problema resulta, cuando menos, desconcertante y su carácter problemático no desaparecerá a menos que tenga lugar una –por lo demás indeterminada– revolución en aquellas discipinas científicas. Las pioneras entre las propuestas argumentativas del primero de estos grupos serían las de White (1986) y McGinn (1989). Por su parte, propuestas como las de Nagel (1998; 2012) o Strawson (1994/2010: 104) contarían entre las más influyentes dentro de esta segunda línea argumentativa. Una sub-rama de esta segunda línea sostiene que la conciencia está vinculada con procesos mecano-cuánticos que tienen lugar en un nivel anatómico inferior al de las neuronas y que la física tendrá que apuntalar la cuestión general de las explicaciones causales en mecánica cuántica antes de poder dar cuenta del problema de la conciencia (Penrose, 1989).

[5] Cerremos este exordio con un más que justificado excurso en nota al pie destinado a hacer justicia a un orden de prioridad que parece haber pasado totalmente inadvertido. La línea argumentativa explotada por Jackson (1982; 1986) y Nagel (1974) había sido desarrollada antes de ellos por el filósofo británico Nicholas Maxwell en dos artículos publicados en la segunda mitad de los sesenta (Maxwell, 1966: especialmente 303-308; 1968: especialmente 127, 134-137, 140-141). Tal y como Maxwell explicara recientemente (Maxwell, 2011: 3), Frank Jackson ha admitido haber leído el segundo de los referidos artículos antes de publicar su argumento epistemológico, mientras que Thomas Nagel, en carta a Maxwell, reconoce la prioridad de éste subrayando la injusticia que supone esta omisión en la historia oficial de los *Consciousness Studies* y afirma, por otra parte, no haber tenido conocimiento de la existencia de dichos artícu-

*El argumento epistemológico*

El argumento epistemológico fue propuesto por Frank Jackson a principios de los ochenta en su célebre artículo "Epiphenomenl qualia". Dicho argumento consiste en un *Gedankenexperiment* que invita al lector a considerar la extraña historia de Mary, una experta en neurofisiología de la visión en color que podría suscribir las palabras de Knut Nordby, investigador acromatópsico de la percepción visual en la Universidad de Oslo:

> Although I have acquired a thorough theoretical knowledge of the physics of colours and the physiology of the colour receptor mechanisms, nothing of this can help me to understand the true nature of colours (Nordby, 1990: 305).

Mary conoce absolutamente todos los detalles de la física y la fisiología de la visión en color –en el escenario descrito por Jackson se nos pide, de hecho, que imaginemos una erudita aun mayor, pues Mary "sabe todo lo que hay que saber acerca de la naturaleza física del mundo" (Jackson, 1986: 291τ)–, pero ha estudiado y vivido todo lo que ha estudiado y vivido en blanco y negro, encerrada en un cuarto en el que todo estaba dispuesto para que no pudiera tener experiencias cromáticas –más allá del blanco y negro–. Mas, he aquí que, prosigue Jackson, al abandonar Mary un feliz día su cautiverio acromático y ver objetos coloreados por vez primera, aprende algo: cómo se ven los colores.[6] La conclusión de Jackson ante este ficticio aprendizaje es la de que el fisicalismo es falso. Parece una conclusión un tanto radical y precipitada. ¿En qué se basa Jackson para defender que esta conclusión se sigue del escenario propuesto? Jackson define casi de pasada el fisicalismo como una tesis según la cual toda la información (correcta) es información sobre fenómenos físicos (Jackson, 1982: 127) o,

los hasta el momento en que Maxwell le puso al corriente de ella. Dada esta habitual laguna, consideramos justo hacer breve mención del modo en que Maxwell llamara la atención sobre el mismo punto que Nagel y Jackson. El núcleo de su argumento consiste en la constatación de que existen unas características reales del mundo (a las que denomina "cualidades sensoriales tal y como las experimentamos") que son tales que, si F es una de ellas, es necesario haber experimentado subjetivamente una clase específica de sensación-F para saber lo que F es. Pero ninguna propiedad física cumple esta definición, y por tanto la física no puede predecir las características de tipo F, es decir, las "cualidades sensoriales tal y como las experimentamos", esto es, los colores, los sonidos o los olores, es decir, los *qualia*.

[6] Mary y el arcángel matemático de Broad (1925) son, poco más o menos, el mismo personaje.

alternativamente, como la tesis de que el mundo en que vivimos es enteramente físico (Jackson, 1986: 291). Tanto las teorías de la identidad de los años cincuenta como los subsiguientes funcionalismos caen dentro de lo que Jackson denomina fisicalismo, y asimismo lo haría cualquier explicación neurobiológica o cognitiva de la conciencia fenoménica. Y bien, el fisicalismo ha de ser falso, según la inferencia que Jackson realiza partiendo de su escenario sci-fi, dado que no es *todo* lo que hay que saber, pues Mary aprende algo –adquiere conocimiento– cuando sale de su cautiverio y ve por primera vez objetos coloreados, y lo que aprende es el modo en que se ven los colores, sus *qualia*. Mary poseía toda la información física en su habitación en blanco y negro, pero su conocimitno era, según Jackson, incompleto. El núcleo de la argumentación consiste así en un intuitivo intento de probar que alguien que poseyera toda la información física no poseería toda la información que cabría poseer, porque ese alguien seguiría sin saber nada acerca de los *qualia*.

En la segunda parte de "Epiphenomenl qualia", Jackson compara su argumento con la que Dennett (2001) denomina "corazonada zombi", ese dispositivo retórico según el cual un mundo poblado por seres física y funcionalmente idénticos a nosotros pero completamente inconscientes es concebible y por tanto posible, de donde, supuestamente, cabe igualmente colegir la falsedad del fisicalismo. La corazonada zombi es, sin embargo, harina de otro costal: hablaremos de ella en el siguiente subapartado. Será suficiente señalar en este punto que Jackson presenta el argumento epistemológico como más potente que la corazonada zombi en tanto no necesita apelar a mundos posibles y no depende, así, de ninguna discutible intuición modal.

De igual modo, en la tercera parte de "Epiphenomenl qualia" Jackson pretende distanciarse de Nagel y su argumento del punto de vista –que enseguida comentaremos– aduciendo que el mismo no consigue refutar el fisicalismo: el argumento de Nagel se basa en nuestra capacidad para imaginar la experiencia de otra criatura y el fisicalismo, según Jackson, no afirma nada acerca de semejante capacidad. Tratando de subrayar la diferencia entre su argumento y el de Nagel, Jackson hace hincapié en una intuición según la cual, si la tesis fisicalista fuera cierta, suficiente información física sobre un sujeto dado habría de ser suficiente para saber todo lo que cabe saber acerca de sus experiencias cromáticas *sin necesidad de ejercitar nuestra imaginación*. Sea como fuere, la verdad es que puede que exista una sutil diferencia entre ambos argumentos, pero, como enseguida comprobará el lector, en el fondo, ambos hacen lo mismo del mismo modo: atacan la idea de que la conciencia fenoménica pueda ser explicada

científicamente sugiriendo que cualquier cantidad de datos científicamente manejables será siempre insuficiente, dado que podremos disponer de los datos y seguir sin saber nada acerca de la conciencia fenoménica.

*El argumento del punto de vista*

El argumento del punto de vista de Nagel ha sido, quizá, el más influyente[7] en la filosofía de la mente contemporánea relacionada con los *Consciousness Studies*, pues ya los propios términos en los que Nagel lo propusiera en el clásico artículo "What is it like to be a bat?", publicado en 1974, puede decirse que resultan omnipresentes e ineludibles en la discusión acerca del aspecto cualitativo o fenoménico de los estados mentales conscientes. Es, pues, casi imposible encontrar un libro o artículo sobre el problema de la conciencia en el que no salgan de un modo u otro a relucir ideas y expresiones tomadas de la formulación original de este argumento. Verdaderos ríos de tinta han corrido por esta vertiente y, curiosamente –curiosamente porque la palabra «*qualia*» no aparece en el referido artículo de Nagel–, buena parte del caudal concierne al debate en torno a los *qualia*. A pesar de que Nagel no mencione en a los *qualia* este artículo, el mismo constituye un verdadero hito dentro de la historia de la cuestión de los *qualia*, dado que en él defiende Nagel la existencia de un peculiar aspecto subjetivo de los estados mentales conscientes ofreciendo a los teóricos posteriores un marco para la discusión de la conciencia fenoménica que ha gozado de gran éxito. La idea que se halla en el núcleo de dicho marco es la de que un organismo tiene o es sujeto de estados mentales conscientes si y sólo si hay algo que es como ser el organismo en cuestión y estar en dichos estados, si y sólo si hay algo que sea para el organismo como tener o estar en esos estados. Esta idea resulta fundamental en el actual debate acerca de la conciencia fenoménica y los *qualia*, dado que existe un acuerdo generalizado según el cual éstos son propiedades de ciertos estados mentales constituidas por cómo es (what it is like) tener o estar en esos estados.

El argumento de Nagel se basa en una asunción según la cual por mucha información que acumulemos y articulemos en el marco de cualquier clase de teoría acerca del sistema perceptivo del murciélago, nunca sabremos qué se experimenta al ser un murciélago y disfrutar de su peculiar rango de modalidades sensoriales; nunca

[7] Dennett (1991a: 441) así lo nota.

sabremos, por ejemplo, cómo es orientarse mediante ecolocalización, porque carecemos de la pertinente modalidad sensorial: nunca hemos experimentado qué se siente al ser un murciélago y orientarse mediante ecolocalización y nunca podremos experimentar cómo es tener esa clase de experiencia porque no somos el tipo adecuado de criatura para ello –nuestra fisiología y nuestra filogenia nos prohíben esa posibilidad–. De la argumentación de Nagel, hemos de entender, se sigue que la experiencia tiene un peculiar carácter subjetivo (el famoso *what it is like*) en virtud del cual cabe dudar de la adecuación de cualquier teoría fisicalista de la conciencia. Nuevamente «fisicalismo» se predica por igual de cualquier clase de aproximación científica a la experiencia consciente, pues aunque la argumentación de Nagel parezca centrarse en aspectos más directamente vinculados con el nivel de implementación –haciendo uso de la famosa distinción de Marr (1982)– que con el algorítmico o el computacional, la conclusión se aplicaría idénticamente a éstos, dado que tanto en el artículo que estamos comentamos (Nagel, 1974: 437) como en escritos posteriores, Nagel defiende que el carácter intrínseco de la experiencia consciente permanece fuera del alcance de las explicaciones funcionales (Nagel, 2000: 433 y ss.). Así, todo intento de solucionar el problema de la conciencia estaría abocado al fracaso, a topar con una barrera infranqueable para la explicación científica.

Nagel presenta su argumento como un obstáculo para cualquier teoría reduccionista de la mente, pero el mismo ha de ser de hecho entendido como un argumento misteriano por cuanto al hablar de reduccionismo Nagel tiene en mente una concepción según la cual todo fenómeno mental puede ser explicado por las ciencias naturales, y por cuanto propone que siempre habrá un elemento esencial de la mente que cualquier explicación naturalista de la misma dejará necesariamente intocado: el carácter subjetivo de la experiencia consciente. Según Nagel, todos los que denomina "análisis reductivos de lo mental" son perfectamente compatibles con la ausencia de este carácter subjetivo de la experiencia consciente, idea con la cual nos hallamos ciertamente cerca de la de los zombis filosóficos –que tardaría aún dos décadas en irrumpir en el debate sobre la conciencia–, pues que los análisis reductivos a los que alude Nagel sean lógicamente compatibles con la ausencia de conciencia fenoménica significa que podemos concebir una criatura instanciándolos a ellos pero no a ella.

Eso que Nagel denomina carácter subjetivo de la experiencia y que posteriormente ha venido denominándose conciencia fenoménica está directamente relacionado con la idea de punto de vista: todo fenómeno subjetivo está esencialmente vinculado a un

punto de vista concreto que, según Nagel, restará necesariamente más allá de las posibilidades explicativas de cualquier teoría fisicalista de la mente, pues si el fisicalismo es cierto, debiera poder ofrecerse una explicación de las propiedades fenoménicas de los estados mentales en términos físicos, pero, según Nagel (1974: 437), cuando examinamos el carácter subjetivo de los mismos, tal posibilidad parece cerrarse. El famoso ejemplo del murciélago viene a ilustrar esta idea: a Nagel no se le ocurre modo alguno en que la potencia explicativa de una teoría fisicalista de la mente pudiera alcanzar a ofrecernos la posibilidad de conocer la experiencia consciente de un murciélago, porque conocerla implica conocer qué es como (what it is like) ser un murciélago, es decir, compartir su punto de vista, posibilidad que, como indicábamos, nuestra fisiología nos prohíbe –aquí, "punto de vista" se refiere a la peculiar forma de experiencia propia de una determinada clase de criaturas, y la forma de la experiencia, es decir, el tipo de punto de vista de una criatura que se orienta mediante ecolocalización nos está vetado de un modo similar a ése en que le está vetado al ciego el mundo fenomenológico del color–. De cara a conocer la naturaleza de la experiencia fenoménica de un murciélago debiéramos poder compartir o encarnar su punto de vista. Esta idea de la inevitabilidad e inexpugnabilidad del punto de vista aparece en la argumentación de Nagel contrapuesta a la de la objetividad: a ésta tienden, huyendo de puntos de vista particulares, las explicaciones reduccionistas en su búsqueda de la verdadera naturaleza de las cosas, pero estando la naturaleza de la experiencia consciente esencialmente vinculada con la subjetividad de un particular punto de vista, parece que el avance reduccionista hacia la objetividad no podrá ayudarnos a aproximarnos al conocimiento de la verdadera naturaleza de la experiencia, de donde extrae Nagel la conclusión de que la experiencia será siempre irreductible.[8]

Nagel hace gala en su argumentación de un más que comprensible –dadas las intenciones de la misma– realismo experiencial, un realismo al que vincula la idea

[8] Nagel ha continuado defendiendo esta postura antirreduccionista en textos más o menos comentados hasta que, recientemente, ha causado cierto revuelo dentro y fuera del ámbito de la filosofía de la mente y las ciencias cognitivas al dar forma a un nuevo hombre de paja (Diéguez Lucena, 2013: 346) y postular la necesidad de abandonar la ortodoxia naturalista e introducir en nuestra concepción del universo unas misteriosas e inespecíficas leyes naturales teleológicas (Nagel, 2012: 72-73, 95-102, 127-131) que den cuenta de aquello que, según Nagel, no puede explicar la ortodoxia materialista neodarwiniana: la vida, la conciencia y la razón. Nagel acompaña su invitación al abandono de la ortodoxia de una propuesta positiva, pero un tanto vaga: con la introducción de esa misteriosa teleología natural, la vida, la conciencia y la razón aparecerían integradas en el marco de una ontología neohegeliana (Ibíd.: 25) que las presenta no como casuales sub-productos de procesos mecánicos obedeciendo leyes ciegas, sino como elementos básicos de la realidad (Ibíd.: 39). Con este libro culminaría, en definitiva, un extraño giro de Nagel desde –en los términos que introdujéramos en la nota al pie cuarta de este capítulo (vid. supra)– su inicial misterianismo *irrestricto* hacia una forma *expectante* de misterianismo.

wittgensteiniana de las fronteras de nuestro lenguaje: dentro de ese reino incontestablemente real de fenómenos mentales existen hechos que caen más allá del alcance de nuestros conceptos; en particular, los hechos que tienen que ver con encarnar un punto de vista concreto y –en términos de la filosofía de la mente posterior a este artículo clásico– ser fenoménicamente consciente.

*b) La brecha explicativa*

La noción de "brecha explicativa" (*explanatory gap*) fue acuñada por Joseph Levine a mediados de los ochenta (Levine, 1983) y perfilada en la década subsiguiente por el mismo autor (Levine, 1993; 1999). La argumentación de Levine consistió inicialmente en una variación sobre el tema planteado en los setenta por Saul Kripke acerca del modo en que los enunciados de identidad entre términos relativos a fenómenos mentales y términos relativos a fenómenos físicos se nos presentan con un aire de contingencia que no suscitan enunciados de identidad entre términos relativos a fenómenos físicos. La intención de Levine, en cualquier caso, era la de elaborar la intuición de que algo queda fuera cuando se formulan identidades entre enunciados referidos a estados o procesos neurofisiológicos y enunciados referidos a estados o procesos fenoménicamente conscientes (las así llamadas identidades psicofísicas). Ya en el segundo párrafo de su seminal artículo de 1983, Levine señala que su acometida antimaterialista es más débil que la de Kripke, pues pretende transformar la argumentación metafísica de éste en una argumentación epistemológica (Levine, 1983: 354). El núcleo de esta argumentación epistemológica consiste en señalar que los enunciados de identidad referidos a fenómenos físicos (agua=$H_2O$; calor=movimiento molecular) no dejan nada fuera y son totalmente explicativos mientras las identidades psicofísicas parecen dejar algo crucial inexplicado (Ibíd.: 357). Según Levine, las identidades que no involucran conceptos referidos a eventos fenoménicamente conscientes son totalmente explicativas en el sentido de que, una vez definimos, por ejemplo, el calor en términos de movimiento molecular, comprendemos perfectamente el modo en que el calor puede jugar los roles causales que efectivamente juega, y una vez entendemos cómo juega estos roles causales no hay nada más que entender al respecto. Por el contrario, cuando definimos el dolor en términos conductuales y neurofisiológicos, sus roles causales pueden entenderse como perfectamente capturados por nuestra definición, pero hay algo que queda fuera de dicha definición y para lo cual la misma no resulta explicativa: el modo en que sentimos

dolor. Al descubrir los mecanismos neurofisiológicos que subyacen a los roles causales que juega el dolor logramos explicar un importante aspecto del mismo, pero el dolor, según Levine, no se agota en sus roles causales y por tanto queda aún algo por explicar: el aspecto fenoménico del mismo. De este modo, concluye Levine, ante cualquier enunciado de identidad entre un evento mental fenoménicamente consciente y un conjunto de fenómenos neurofisiológicos, nada en el lado neurofisiológico del enunciado permite entender el modo en que en éste encaja el lado fenomenológico de la identidad predicada en el mismo. Así las cosas, Levine se siente en posición de afirmar que la conexión entre ambas partes de un enunciado de identidad de este tipo resultará siempre completamente misteriosa.

Levine salta en su argumentación del comentario acerca del modo en que nos sentimos cuando somos expuestos a enunciados de identidad a la especulación acerca de lo explicable e inexplicable en un intento por convencernos de que cualquier explicación fisicalista de la conciencia fenoménica dejará fuera, precisamente, la propia conciencia fenoménica. Su intento se agota en aseveraciones que enuncia con una claridad y un candor cuestionables en sentencias del tipo de la siguiente: no hay nada en cualquiera que sea el conjunto de procesos neurofisiológicos del que pretendamos derivar una explicación de un evento mental fenoménicamente consciente que explique o haga inteligible el peculiar carácter subjetivo de ese evento consciente. Llamemos a ese conjunto de procesos X y supongamos que son ellos la base de la explicación neurofisiológica que pretendemos elaborar para el dolor de muelas. Lo que Levine afirma es que no hay nada que podamos determinar acerca de X que explique por qué instanciar X ha de tener el carácter cualitativo que tiene un dolor de muelas, que lo que sentimos cuando instanciamos X –si X contara como descripción correcta de la neurofisiología de un dolor de muelas– ni queda explicado ni es inteligible a la luz de nuestra completa comprensión de las propiedades funcionales o físicas de X.

Levine postula de este modo la ininteligibilidad del nexo entre un evento mental fenoménicamente consciente y su "correlato físico subyacente". De aquí partiría la joven tradición que en la filosofía de la mente contemporánea ha conceptualizado la brecha explicativa como una supuesta capacidad humana para derivar sin perplejidad o remilgo epistémico de ninguna clase consecuencias acerca de cualesquiera hechos físicos a partir de información acerca de condiciones iniciales y leyes científicas, acompañada de una supuesta incapacidad para hacer lo mismo cuando se trata de derivar consecuencias acerca de acontecimientos fenoménicamente conscientes a partir de informa-

ción acerca de condiciones iniciales neurobiológicas y leyes científicas relativas al funcionamiento del sistema nervioso. Así, según la forma en que acabara extendiéndose en los *Consciousness Studies* la intuición de la existencia de una brecha explicativa en nuestra comprensión del modo en que la actividad neurofisiológica deviene experiencia consciente, incluso comprendiendo detalladamente todo lo que las ciencias de la mente, el cerebro y la conducta puedan decirnos acerca de la relación entre el sistema nervioso y la mente consciente, esa relación continuará siendo un misterio (Papineau, 2011). Según esta intuición, pues, teniendo presentes los hechos y leyes científicas relevantes, concluir que un volumen dado de agua hervirá al alcanzar los 212ºF a una atmósfera resulta epistémicamente aproblemático y no causa ninguna perplejidad (Levine, 1999). Lo contrario sucedería, en cambio, con cualquier cantidad de hechos y leyes científicas que pretendamos utilizar en la explicación de un evento consciente. Siempre podríamos concebir la historia causal que esos hechos y leyes desplieguen ante nosotros sin que en la misma hiciera acto de presencia evento fenoménicamente consciente alguno. Levine y el resto de los autores que comparten su intuición misteriana creen que la misma les permite aseverar que conociendo todos los hechos y leyes relevantes, no podríamos derivar hechos acerca de la conciencia fenoménica del modo en que, supuestamente, podemos hacerlo con cualquier propiedad física de cualquier sistema. De este modo, presumen estos misterianos, disponiendo de todos los datos relevantes, podríamos derivar *a priori*, *válida* y *aproblemáticamente* –sin perplejidad alguna– cualquier consecuencia acerca de cualquier sistema físico ("explotará", "hervirá", "un campo magnético no podría penetrarlo", "si lo enfrío mucho será un superconductor", etc.), pero no sucedería lo mismo en el caso de que ese sistema físico sea consciente y lo que tratemos de derivar sean consecuencias acerca de sus estados mentales fenoménicamente conscientes. Ninguna cantidad de conocimientos científicos acerca de dicho sistema nos permitirá derivar a priori hechos acerca de su mundo fenomenológico partiendo de hechos acerca de su constitución física. En este sentido, Levine (2007: 377) afirma que podríamos conocer *todo* acerca del modo en que una criatura se halla constituida y aún así continuar preguntándonos si es o no consciente.

Según Levine, el argumento de la brecha explicativa prueba que existe una brecha epistemológica, pero no una ontológica. Esto vendría a significar que no acabaremos de encontrarnos satisfechos con ninguna explicación científica de la experiencia consciente –incluso aunque fuera correcta–, que siempre podríamos concebir los fenómenos y relaciones a los que aluda la explicación acompañados de cualquier tipo de experiencia

consciente o de ninguna en absoluto, mientras que lo mismo no sucedería con ninguna otra clase de explicación científica. A diferencia de otros misterianos, Levine no defiende que la conciencia fenoménica no encaje en el orden de los fenómenos naturales, ni que la ontología fisicalista sea inadecuada. Lo que propone es que, aunque dicha ontología sea correcta, aunque la conciencia fenoménica se halle enteramente constituida por fenómenos físicos, no tenemos la menor idea de cómo explicar que determinados conglomerados de fenómenos físicos sean conscientes ni podríamos en ningún caso entender cómo encaja la conciencia en el orden de los fenómenos naturales.

*c) La clausura cognitiva*

La tesis de la clausura cognitiva ha sido elaborada y defendida por Colin McGinn en sucesivas publicaciones desde su primera formulación, a finales de la década de los ochenta (McGinn, 1989). No obstante, con la tesis de la clausura cognitiva de McGinn no vio la luz una idea nueva, sino una vieja remozada, rebautizada y adornada con los atavíos propios de la filosofía de la mente de corte analítico.

> The passage from the physics of the brain to the corresponding facts of consciousness is unthinkable. Granted that a definite thought, and a definite molecular action in the brain occur simultaneously; we do not possess the intellectual organ, nor apparently any rudiment of the organ, which would enable us to pass, by a process of reasoning, from the one to the other. (Tyndall, 1871: 119-120)

El pasaje citado, a pesar de resultar bastante actual, pertenece al discurso presidencial que John Tyndall[9] pronunciara en Norwich, en 1868, para la *Physical Section of the British Association for the Advancement of Science* (Russel Wallace, 1870: 163). Tyndall no buscó una forma de referirse a esta idea, pero si McGinn hubiera tenido que hacerlo por él, sin ninguna duda la habría bautizado como "tesis de la clausura cognitiva". También el filósofo y economista liberal austro-británico Friedrich August von Hayek (Nobel de Economía en 1974) se adelantó a McGinn cuando en su ensayo de 1952 sobre los fundamentos de la psicología (en el que desarrollara en un plano teórico un marco hebbiano[10] para el análisis de la memoria y el aprendizaje) dejó dicho que la

[9] "El físico irlandés que logró explicar el color azul del cielo" (Koch, 2012: 25τ), según las palabras que Cristof Koch le dedica al citar este mismo fragmento en su *Consciousness. Confessions of a Romantic Reductionist*.

[10] A pesar de que la obra en la que Hebb expone su famoso modelo del refuerzo de las conexiones sinápticas (Hebb, 1949) –habitualmente denominado *regla de Hebb*, *postulado de aprendizaje de Hebb* o *teoría de la asamblea celular*– apareció publicada tres años antes que la obra de Hayek a la que aludimos,

idea de la mente comprendiéndose a sí misma es un contrasentido (Hayek, 1952: 192). No obstante, y a pesar de estos antecedentes más o menos explícitos, ha sido sin duda McGinn quien más páginas ha dedicado a la elaboración, defensa y difusión de esta tesis, y será por tanto a su planteamiento al que atenderemos en nuestra exposición de la misma.

Las conclusiones misterianas de McGinn no se restringen a la conciencia o la filosofía de la mente, sino que abarcan la agenda filosófica occidental al completo, incluyendo, claro, problemas tradicionales como el de la libertad, el del significado o el del conocimiento (McGinn, 1993). Aquí nos referiremos sólo de pasada a su planteamiento misteriano general por cuanto el mismo constituye el contexto en que se enmarca su misterianismo respecto de la conciencia. En la propuesta misteriana general de McGinn, los irresolubles misterios a los que la filosofía se enfrenta se alzan como tales a causa de nuestros procedimientos epistémicos, dado que, según McGinn, el éxito en las áreas del conocimiento que se han mostrado más vigorosas y eficaces (como la física) habría provenido de estrategias explicativas bottom-up gracias a las cuales se habrían desarrollado teorías agregativas sucesivamente capaces de dar cuenta de un rango mayor de fenómenos.[11] Así, siempre según McGinn, el problema estribaría en que para los fenómenos a los que los filósofos destinan vanos esfuerzos no existe la posibilidad de recurrir a un procedimiento explicativo análogo, dado que entre los fenómenos básicos desentrañados mediante aquellas estrategias explicativas bottom-up y los fenómenos que aborda la filosofía existen significativos –insalvables, de hecho– hiatos (McGinn, 2002: 209), de donde concluye –extraña y heroica conclusión para un hombre que ha dedicado su vida a la filosofía, dicho sea tangencialmente– que la filosofía no sería más que una ocupación fútil (Ibíd. 210).

El misterianimo de McGinn respecto de la conciencia, al igual que su propuesta misteriana general, es un misterianismo, como el de Levine, de corte epistemológico, pues entiende el filósofo británico que –según la paráfrasis de Dennett (1991b) en su reseña de McGinn (1991)– del hecho de que nosotros no podamos desentrañar el problema de la conciencia no debemos extraer la conclusión de que ésta sea intrínsecamente misteriosa. La tesis de la clausura cognitiva, el núcleo del misterianismo de McGinn

ésta es fruto de una idea del austro-británico anterior incluso a sus estudios de economía, una idea que aplicara treinta y dos años después de que le asaltara por vez primera las mientes a una teoría global del funcionamiento cerebral, el aprendizaje y la memoria.

[11] Esta propuesta epistemológica corre en paralelo a la ontología que McGinn defiende, una ontología en la que la naturaleza toda aparece constituida como un sistema de entidades agregadas y derivadas en el que las menos básicas se hallan conformadas por las más básicas (McGinn, 2002).

respecto de la conciencia, propone que el problema de la conciencia es antes un misterio –en sentido chomskiano– que un problema porque nuestras capacidades cognitivas son insuficientes para desentrañarlo de igual modo que las capacidades cognitivas de un mono tití son insuficientes para comprender el teorema de Bayes.[12] McGinn cree encontrarse pues a suficiente *altura histórica*[13] como para aseverar que la reluctancia del problema de la conciencia, antes que deberse al provisorio grado actual de desarrollo empírico, tecnológico, metodológico, conceptual y teórico en las disciplinas científicas pertinentes, es de raigambre biológica: al igual que Nagel propuso que no somos el tipo de criatura apropiado para experimentar la modalidad sensorial de la ecolocalización, McGinn propone que no somos el tipo de criatura apropiado para resolver el problema de la conciencia, esto es, *para nosotros*, y por lo tanto, el *misterio* de la conciencia. Quizá mañana llegue a la Tierra una nave espacial –procedente, verbigracia, de algún lugar de la Galaxia del Cigarro (a.k.a. Meiser 82)– pilotada por criaturas capaces de explicarnos la solución al problema de la conciencia, pero si McGinn está en lo cierto, no podríamos entender una sola palabra. Así pues, aunque la conciencia y su vínculo con los fenómenos biológicos oportunos puedan articularse en una teoría verdadera, nosotros ni podemos formularla ni podríamos comprenderla. McGinn, en definitiva, creía hace ya más de un cuarto de siglo hallarse a la suficiente altura histórica como para asegurar que había llegado el momento de admitir con franqueza que los seres humanos no podremos desentrañar jamás el misterio de la conciencia (McGinn, 1989: 349).

La tesis de la clausura cognitiva no conduce a McGinn a conclusiones dualistas, pues está convencido de que "la conciencia es un fenómeno natural indisolublemente ligado al cerebro, pero (…) que constituye una propiedad incognoscible de este órgano para este órgano" (Díaz, 2008: 239). En lugar de ello, McGinn, como apuntábamos, defiende una ontología naturalista a la que suma un planteamiento epistemológico misteriano según el cual la conciencia se encuentra más allá de una de las fronteras impuestas al conocimiento humano y, por lo tanto, inaccesible a los métodos de las ciencias naturales. Según Dennett (1991b), McGinn estaría apoyándose en este punto en Fodor –

[12] McGinn ofrece una definición "formal" de su noción de clausura cognitiva. Esta definición "formal" no es imprescindible –y puede que tampoco de gran ayuda– para comprender la tesis, motivo por el cual la incluimos sólo en nota al pie: "A type of mind M is cognitively closed with respect to a property P (or theory T) if and only if the concept-forming procedures at M's disposal cannot extend to a grasp of P (or an understanding of T)" (McGinn, 1989: 350; 1991: 3).

[13] Este giro de corte orteguiano pretende aludir a la historia de la ciencia y, particularmente, a la de las jóvenes disciplinas encargadas de la mente, la conducta y el sistema nervioso.

que a su vez se habría apoyado en la citada intuición categórica de Chomsky (1976)– a la hora de formular y defender su tesis de la clausura cognitiva, y en particular, en la noción de *epistemic boundedness* (Fodor, 1983), según la cual existen restricciones endógenamente determinadas para los tipos de problemas que los seres humanos podemos resolver.[14] Uno de los motivos por los que entiende que esas restricciones obliteran ineluctablemente la vía hacia la explicación de la conciencia tiene que ver con que –más allá del hecho de que, como indicábamos, no se hallen disponibles en el caso de la conciencia, como tampoco en el del resto de los problemas filosóficos, las referidas estrategias explicativas bottom-up y teorías agregativas– el medio por el cual accedemos al mundo de la conciencia y el medio por el cual accedemos al conocimiento neurocientífico guardan un mutuo silencio: ni la introspección parece expresar nada acerca de la neurofisiología ni viceversa, ni, decisivamente, ninguna de ambas parece ofrecer pista alguna acerca de su vínculo. Tanto mediante introspección como mediante cualesquiera métodos neurocientíficos, la naturaleza del vínculo entre fenomenología y fisiología se mostrará, según McGinn, renuente.

El problema que McGinn juzga insoluble es pues el de esclarecer el modo en que la experiencia consciente surge de la actividad neuronal, esto es, de procesos netamente físicos. McGinn, en este sentido, entiende que ese surgir ha de suceder en virtud de una propiedad natural del cerebro (a la que denomina P), esto es, una propiedad que no excedería el marco de aquellos procesos netamente físicos y que podría ser objeto de una teoría científica capaz de explicar cabalmente el modo en que la actividad neurofisiológica da lugar a la experiencia consciente. Esta propiedad P juega en la argumentación de McGinn el rol de vínculo psicofísico y es presentada como una propiedad natural, entendiendo por natural "de algún modo instanciada en el mundo físico". Como adelantábamos, McGinn opina que estamos cognitivamente cerrados a esta propiedad, que nos es inaccesible; pero en su argumentación, clausura cognitiva respecto de P no significa irrealismo respecto de P. Esto es, McGinn sostiene que dicha propiedad existe y puede ser adecuadamente integrada en una teoría científica, pero añade que nosotros, a causa de las limitaciones de nuestro aparato cognitivo, no somos el tipo de criatura capaz de formular o comprender esa teoría. Disponemos, ciertamente, de la posibilidad de acceder a los dos extremos del binomio: podemos acceder directamente a la experiencia consciente mediante introspección y podemos investigar el funcionamiento del cerebro

[14] Fodor, Nagel y McGinn constituyen el núcleo de la que Dennett, irónicamente, denomina "escuela nihilista de New Jersey".

desde el nivel molecular al anatómico mediante las diferentes técnicas disponibles en neurociencias. No obstante, como apuntábamos, cada una de estas dos caras de la moneda parece guardar un circunspecto silencio tanto acerca de la otra como, crucialmente, acerca de su interrelación, con lo cual, esa propiedad natural vinculante postulada por McGinn permanece inaccesible para nosotros.[15] Siendo esa propiedad P una propiedad natural pasible de incardinación en una teoría capaz de explicar adecuadamente el modo en que con ella el agua de la fisiología se convierte en el vino de la fenomenología, ¿por qué entiende McGinn que no podremos dar con ella? McGinn comienza por responder apelando al hecho de que los conceptos que podemos formar partiendo de nuestra experiencia consciente dependen de las modalidades de experiencia de las que somos capaces –así, un ciego no puede captar en puridad un concepto visual, como tampoco podría ningún ser humano comprender cabalmente un concepto formado a partir de experiencias de la modalidad sensorial de la ecolocalización de que disfruta un murciélago–. De forma que existen unas restricciones fenoménicas que limitan nuestra capacidad de formar conceptos, y McGinn argumenta que dicha limitación juega en contra de nuestras posibilidades de dar con la señalada propiedad P. Su argumento depende de la asunción según la cual la comprensión de la propiedad física que se halle a la base de una experiencia subjetiva específica, también conferirá comprensión del carácter subjetivo de esa experiencia. El razonamiento de McGinn es el siguiente: presupone que el conocimiento de la referida propiedad P habría de conferir también una comprensión de la teoría que deba dar cuenta del modo en que esa propiedad traza el vínculo entre fisiología y fenomenología, es decir, de la teoría que, *ex hypothesi*, explica cómo la conciencia depende de aquella propiedad P. Pero para entender esa teoría tendríamos asimismo que entender los términos en los que se halle formulada, y uno de ellos, según McGinn, será el de la cualidad subjetiva de un estado cerebral, y entender completamente una cualidad subjetiva significa captar su carácter, de donde infiere McGinn que conocer P nos conferiría un conocimiento que no podemos tener dados los límites fenoménicos de nuestra capacidad de formar conceptos, y así, concluye, por *reductio ad absurdum*, no podemos llegar a conocer P.

El siguiente segmento de la respuesta de McGinn a la pregunta acerca de los motivos por los cuales entiende que no podremos dar con P reza así: porque no se trata de una propiedad perceptible ni de una propiedad que pueda derivarse de propiedades per-

[15] "No matter how many ways you try to arrange the constituents of the brain you never end up with a conscious experience" (McGinn, 1982/1996: 46).

ceptibles. Ni la conciencia ni, mucho menos, el vínculo entre ella y la fisiología pueden observarse "mirando el cerebro": P está también perceptivamente cerrada (*perceptually closed*) para nosotros. Si esta respuesta puede sonar ya vaga, poco cabe decir de los motivos que llevan a McGinn a formularla, pues según sus intuiciones –y suponiendo la clara referencialidad de los términos en los que las expresa–, dicha propiedad ha de ser no-perceptible porque no podemos imaginar nada perceptible en nuestra neurofisiología que haga comprensible el modo en que de ésta surge la experiencia consciente: he aquí un primer devaneo de McGinn con la estrategia argumentativa "concebibilista", una estrategia endémica en la filosofía de la mente de las últimas décadas que describiremos brevemente en el próximo subapartado y criticaremos en el siguiente apartado. En el mundo neuronal todo se nos presenta, además, espacialmente, pero McGinn encuentra que ningún percepto espacial puede ofrecernos las claves del vínculo entre la conciencia y la neurofisiología: no daremos –vaticina– con la naturaleza del vínculo entre fenomenología y fisiología apoyándonos en las propiedades espaciales que percibimos estudiando el cerebro. No pudiendo percibir P, podríamos, sin embargo, alcanzar conocimiento de P mediante inferencia a la mejor explicación desde lo que efectivamente percibimos. En este punto McGinn da el siguiente paso en su argumentación al proponer que nuestra forma de elaborar conceptos se muestra también incapaz de aproximarnos a la elucidación de esa propiedad natural P, vínculo entre fenomenología y neurofisiología, dado que, según un inusitadamente empirista McGinn,[16] el modo en que formamos conceptos científicamente aprovechables se basa en extensiones analógicas de la percepción, pero ninguna extensión analógica de lo que percibimos al estudiar el cerebro podrá conducirnos a esa esquiva propiedad natural P. A lo expuesto añade McGinn, sumándose subrepticia pero ya decididamente a las huestes de la concebibilidad, que la conciencia no parece jugar ningún papel en la explicación de los hechos neurofisiológicos y, por tanto, dado que la conciencia no es necesaria para explicar los datos, aquella propiedad P tampoco lo es.[17] Esta afirmación cumple asimismo la función de probar la imposibilidad de la inferencia a la mejor explicación partiendo de los datos neurocientíficos. Todo lo que tenemos en esos datos se explica sin recurrir a la conciencia, presume McGinn, de forma que las propiedades teóricas que formulemos para explicar esos da-

[16] El adverbio responde al hecho de que el propio McGinn (1989) critica la teoría empirista clásica de la formación de conceptos.

[17] Señalábamos que con este argumento se suma McGinn a las "huestes de la concebibilidad" por cuanto el mismo supone que McGinn ha logrado concebir nada menos que el conjunto completo de hechos y modelos explicativos neurocientíficos y que, además, se ha dado cuenta al concebir semejante plétora de datos y marcos teóricos que en ella la conciencia no juega ningún papel.

tos tampoco requerirán de la inclusión de la conciencia. Resumiendo, estamos cerrados perceptivamente a P, y también lo estamos a cualquier estrategia de formación conceptual e inferencia explicativa –basada en la percepción– que pudiera conducirnos a P: cualquier inferencia a la mejor explicación que ensayáramos partiendo de datos neurofisiológicos se mostraría incapaz sacarnos del mundo *meramente* neurofisiológico en el que la pirotecnia technicolor de la fenomenología no se digna a hacer acto de presencia.

### 2.2._Argumentos misterianos parciales

En este segundo subapartado ofreceremos un breve esquema del partido que unos cuantos filósofos misterianos han pretendido sacar de nuestra supuesta capacidad para concebir una curiosa serie de escenarios hipotéticos. Estos escenarios están diseñados para atacar posiciones funcionalistas y representacionalistas en filosofía de la mente y, en general, en ciencias cognitivas, pues con ellos pretende hacerse intuitivamente evidente que lo fenoménico no puede reducirse a lo funcional, que el contenido representacional no es todo lo que hay que explicar, que los *qualia* no podrían capturarse en una explicación funcional y que, por lo tanto, la conciencia fenoménica estaría más allá de las posibilidades con las que el funcionalismo cuenta de cara a dar adecuada cuenta de los estados mentales y, asimismo y por idénticos motivos, más allá de las posibilidades explicativas de las teorías representacionales y cognitivas de la conciencia.

Antes de empezar a trazar el señalado esquema, maticemos que nos hallamos en realidad ante dos grupos de argumentos: los basados en la concebibilidad (y por tanto, *supuestamente*, la posibilidad) de *qualia* ausentes o zombis, y los basados en la concebibilidad (y por tanto, *supuestamente*, la posibilidad) de espectros invertidos.

#### *a) Los zombis*

Los zombis hicieron una tímida incursión en la filosofía de la mente de la primera mitad de la década de los setenta (vid., v. g., Campbell, 1970; Kirk, 1974a; 1974b),[18] pero no fue hasta la de los noventa que, de la mano de David Chalmers, encontraran en

[18] No está de más señalar que el filósofo y psicólogo británico George Frederick Stout había avanzado ya la idea en las lecciones que impartiera entre 1919 y 1921 en la Universidad de Edinburgo y que aparecerían publicadas diez años más tarde bajo el título *Mind and Matter* (vid. Stout, 1931: 138).

ella su espacio. El término «zombi» es utilizado en filosofía de la mente para hacer referencia a una réplica exacta de un ser humano consciente que, sin embargo, carece de conciencia fenoménica: para mi zombi réplica, hemos de entender, no hay nada que sea como ver una pelota o sentir un pelotazo. No obstante, mi zombi réplica se comporta exactamente igual que yo y manipula información del mismo modo que yo: sus estados mentales tienen idéntico contenido representacional que los míos, pero carecen de aspecto fenoménico o cualitativo –los zombis sólo tendrían pues, en términos de Block (1994a; 1995), conciencia de acceso, y no habría en ellos ni rastro de conciencia fenoménica.

El argumento puede plasmarse en forma silogística como sigue:

Premisa mayor: Los zombis son concebibles.
Premisa menor: Cualquier cosa concebible es posible.
Conclusión: Los zombis son posibles.

Y si los zombis son posibles, según los defensores de esta estrategia argumentativa, entonces habría sido probado que la conciencia fenoménica tiene un carácter autónomo y es radicalmente diversa del contenido intencional o representacional de los estados mentales conscientes. Siendo ese contenido el *explanans* de las teorías cognitivas y representacionales, las mismas habrían sido refutadas en la medida en que el argumento alcance a hacer explícita la imposibilidad de articular cualquier clase de caracterización funcional capaz de incluir a la conciencia fenoménica.[19]

Variables de este tipo de argumento que no comentaremos aquí son los argumentos de la nación China, expuesto en Block (1978), y el comentadísimo argumento de la habitación china (Searle, 1980), los cuales vendrían a ilustrar de diversos modos –y en particular el primero, pues el objetivo del segundo es más ambicioso– que la organización funcional no determina la conciencia fenoménica.

*b) Los espectros invertidos*

[19] Nótese que cabe extender estas conclusiones a una forma radical de misterianismo en un sencillo movimiento: si cabe concebir un zombi físicamente idéntico molécula a molécula, átomo a átomo, muon a muon a un ser humano consciente, entonces, cualquier clase y cantidad de datos manejable por las ciencias naturales es menos de lo que hace falta para elaborar una teoría explicativa de la experiencia consciente. Esta sería la versión molécula-por-molécula del argumento de la concebibilidad de zombis (Chalmers, 1996: cap. 3).

Este segundo grupo de argumentos constituye uno de los más destacados núcleos del ataque contemporáneo a planteamientos funcionalistas en filosofía de la mente y ciencias cognitivas. Aunque la formulación de este tipo de argumento puede remontarse a Locke (1690, lib. II, cap. 32, § 15), aquí nos ocuparemos exclusiva y sumarísimamente del modo en que es esgrimido actualmente. Existen dos variantes: la interpersonal y la intrapersonal. Ambas se nutren de la intuición según la cual cabe concebir variación (inversión) fenoménica manteniéndose constante el contenido intencional –incluso manteniéndose constante la intrincada maraña al completo de relaciones causales, tanto entre estados mentales como entre éstos y reacciones fisiológicas y conductuales–. Así, según la primera versión, dos individuos idénticos conductual y funcionalmente podrían tener experiencias cromáticas sistemáticamente invertidas a pesar de reaccionar *de forma idéntica* ante los mismos estímulos. Por ejemplo, ante la pregunta acerca del color de mi camisa ambos responden "verde", a pesar de que uno la *ve* roja y el otro verde. Este tipo de argumento no se limita a presentar la posibilidad de una conducta verbal idéntica aunada a experiencias subjetivas invertidas: toda otra disposición reactiva, además de la verbal, ha de considerarse que varía sin afectar en absoluto a la presentación de los *qualia* en la conciencia. En la versión intrapersonal, por su parte, la inversión fenoménica se produce *en* el mismo individuo: un 29 de febrero el señor Reversiblez se despierta y, para su sorpresa, comprueba que *ve* de un extraño tono amarillo anaranjado las anteriormente azules paredes de su habitación; abre la ventana y comprueba que el cielo tiene ese mismo tono, y que el resto de los colores le *parecen* diferentes, que sus experiencias cromáticas están, en definitiva, patas arriba. El señor Reversiblez, continúa el argumento, comienza a adaptarse e invierte el nombre de los colores tal y como se presentan en su experiencia subjetiva: dice de la señora Reversiblez que sus ojos son de color verde claro, a pesar de que al mirarlos los *ve* de una suerte de rojo violeta, y hace exactamente lo mismo con el resto de sus experiencias cromáticas y su conducta verbal sobre las mismas –un importante supuesto adicional es el de que, nuevamente, el resto de sus disposiciones reactivas se normalizarían, al igual que su conducta verbal, tras la etapa de adaptación. [20]

[20] Nótese que obtener una variable misteriana radical de esta clase de argumento es tan sencillo como en el caso anterior: sólo hay que extender la idea de que cualquier caracterización funcional es menos de lo

## 3._Crítica de los argumentos misterianos

### 3.1._Por qué no funciona la réplica habitual

La estrategia más común entre los filósofos de la mente fisicalistas interesados en interceptar las conclusiones de los argumentos misterianos ha sido la denominada –desde Stoljar (2005)– *phenomenal concepts strategy* (Balog, 1999; Loar, 1990; Lycan, 1996; Papineau, 1993; Perry, 2001; Sturgeon, 1994; Tye, 1995; 2000). Esta estrategia pretende hacer frente al desafío de los argumentos misterianos partiendo de una base compartida con los proponentes de dichos argumentos: que los escenarios que describen en sus argumentos son concebibles y que de ello cabe extraer alguna clase de conclusión. Así, estos filósofos asumen que hay una forma cabal de interpretar la que denominaremos "jerga de la concebibilidad". La idea que hallamos a la base de dicha jerga es la siguiente: la noción de zombi es completamente coherente y nosotros podemos imaginar sin problema un ser humano molécula a molécula y extraer de ese acto de imaginar conclusiones legítimas; asimismo, que alguien pueda tener en mente todos los datos acerca de la neurofisiología de la visión en color es perfectamente concebible; tampoco nos resulta problemática en absoluto la idea de una mente capaz de manejar todos los datos pertinentes acerca de cualquier sistema físico y, es más, sabemos que tal mente podría derivar a priori de esos datos cualquier consecuencia acerca de dicho sistema y que los resultados de su apriorístico colegir le resultarían totalmente pedestres excepto cuando se trate de deducir hechos acerca de la conciencia fenoménica. Nosotros, quizá a causa de una subdesarrollada capacidad para imaginar, no tenemos muy claro qué cabe pensar de conclusiones alcanzadas partiendo de premisas enunciadas en jerga de la concebibilidad. De este modo, tampoco tenemos muy clara cuál ha sido la utilidad de la enorme cantidad de trabajo argumentativo realizado con vistas a refinar la estrategia de los conceptos fenoménicos (en lo sucesivo PCS, por sus siglas en inglés). Acercarnos a los motivos que han venido impeliendo al uso de esta estrategia nos servirá para delimitar y comenzar a exponer nuestra forma de enfrentar el desafío de los argumentos misterianos.

---

que necesitamos para dar cuenta de la conciencia fenoménica sustituyendo la locución «cualquier caracterización funcional» por «cualquier caracterización física».

¿En qué consiste, pues, la PCS? Según esta línea de argumentación, nuestra posesión de conceptos fenoménicos explicaría los hechos que, según los defensores de posturas misterianas, debieran conducir al abandono de la idea de que la conciencia es un fenómeno tan físico como un riñón y tan explicable como la formación de orina –hechos tales como que parezca intuitivamente obvio que existe una brecha explicativa, que los zombis o los espectros invertidos sean concebibles o que Mary aprenda algo–. ¿Y qué son los conceptos fenoménicos? Los conceptos fenoménicos, se asume, son una clase particular de conceptos cuyo rasgo distintivo estriba en que la cualidad subjetiva de las experiencias de un sujeto dado forma, de algún modo, parte del contenido de los conceptos con los que se refiere a dichas experiencias. Hay que haber tenido experiencias subjetivas como aquéllas a las que apuntan los conceptos fenoménicos para poseerlos e, igualmente, para comprenderlos. Los conceptos fenoménicos son aquéllos que utilizamos en virtud de un modo conceptual de percatación del carácter fenoménico de nuestra experiencia: dicha experiencia tendría en cualquier caso ese carácter, pero, cuando atendemos al mismo, lo subsumimos bajo el concepto fenoménico correspondiente, que podemos entonces utilizar para comunicarnos o, sencillamente, para pensar.

Los filósofos que recurren a la PCS entienden que los conceptos que utilizamos para referirnos a los hechos físicos y los conceptos fenoménicos son diversos y de algún modo incompatibles, dado que los conceptos fenoménicos están, de hecho, aislados del resto de nuestra economía conceptual. Permaneciendo ambas clases de conceptos aisladas, da igual cuánta información en términos de conceptos físicos nos sea proporcionada, da igual que podamos comprender íntegramente *toda* la información acerca del estado físico de un organismo: al no tener los conceptos en los que esa información se halla expresada vínculos con los conceptos mediante los cuales podemos pensar en nuestra experiencia consciente, nada acerca de ésta puede derivarse partiendo de aquella información. Así, la completa heterogeneidad entre lo físico y lo fenoménico que sugieren los escenarios que los argumentos misterianos nos piden que imaginemos se debería no a una heterogeneidad real que ubicaría a las naturalezas de lo físico y lo fenoménico en planos ontológicamente irreconciliables, sino al hecho de que cuando pensamos en una sensación de rojo y en la fisiología de dicha sensación, pensamos en el mismo fenómeno –un fenómeno físico–, pero manejando diferentes clases conceptos: conceptos fenoménicos en un caso y fisiológicos en otro. Son heterogéneas, pues, según esta réplica fisicalista, no las naturalezas de lo físico y lo fenoménico, sino las clases

integradas por los conceptos fenoménicos, de una parte, y los conceptos físicos, de otra. Dada esta heterogeneidad, es comprensible que haya una brecha explicativa, que podamos concebir zombis o que Mary aprenda algo.

Pero, ¿a qué se debe esta heterogeneidad? ¿Qué hay de peculiar en los conceptos fenoménicos que nos impide trazar nexos entre ellos y los físicos? Con la respuesta a esta pregunta rompen filas los defensores de la PCS. Así, por ejemplo, Perry (2001) y O'Dea (2002) sostienen que los conceptos fenoménicos son una clase de conceptos indéxicos (para una introducción a esta noción vid. Perry, 2006). Desde su punto de vista, los conceptos fenoménicos apuntarían a estados neurofisiológicos en una forma indéxica de presentación, de forma análoga a ésa en que el indéxico «yo» apunta a su referente (a mí) cuando digo "yo estoy en Salamanca". Dado que, se asume, conocer todos los hechos no-indéxicos sobre el mundo no permitiría deducir ningún hecho presentado indéxicamente, igualmente, conocer todos los hechos físicos no permitiría deducir ningún hecho fenoménico, lo cual no supone ninguna amenaza para el fisicalismo por cuanto los conceptos fenoménicos, de este modo aislados, pueden perfectamente referir a estados neurofisiológicos en una forma indéxica de presentación aun cuando la más articulada de nuestras redes de conceptos físicos no logren aproximarnos al contenido de aquéllos. Por su parte, Papineau (2002) mantiene que los conceptos fenoménicos son diversos del resto de nuestro aparato conceptual en tanto sólo ellos contienen los estados a los que refieren y, en este sentido, los define como *quotational concepts*. Según su propuesta, los estados mentales a los que se refieren los conceptos fenoménicos se hallan insertos en dichos conceptos. Así, cuando digo "ésta es mi experiencia visual del cielo azul", el concepto fenoménico «ésta» contendría una experiencia visual concreta. En el modelo cuotacional de Papineau los conceptos fenoménicos consisten en un operador de la forma "esta experiencia:___", anexo a una experiencia introspectiva –bien se trate de una experiencia actual o una recreación imaginativa de una experiencia– que, de algún modo, vendría a rellenar el espacio vacío entre los dos puntos y el cierre de las comillas.[21] También desde el punto de vista de esta aproximación a la naturaleza de los conceptos fenoménicos cabe inferir que no

[21] Con posterioridad Papineau (2006) abandonaría el modelo que, calcando desde el inglés (quote), hemos denominado cuotacional –un modelo que explota la analogía entre el modo en que entrecomillar una palabra dirige la atención del lector a la propia palabra (una palabra entrecomillada, desde esta perspectiva, se leería como "esta palabra") y el supuesto modo en que los conceptos fenoménicos contienen *sus* experiencias– sustituyéndolo por otro en el que la individuación de los conceptos fenoménicos depende parcialmente de roles conceptuales. Abundar en el particular, no obstante, proporcionaría sólo ocasión para una no excesivamente justificada digresión. El lector interesado dispone, en cualquier caso, de las fuentes.

podríamos deducir qué conceptos fenoménicos se aplican a algo partiendo de un conocimiento tan completo como quepa imaginar acerca de los conceptos no fenoménicos aplicables a ese algo, lo cual serviría para explicar la intuida existencia de la brecha epistémica entre lo fenoménico y lo físico sin amenazar la ontología fisicalista –dado que, una vez más, el referente de un concepto fenoménico, es decir, el estado que vendría a rellenar el espacio vacío entre los dos puntos y el cierre de las comillas, puede ser un estado neurofisiológico aun cuando nuestros conceptos neurofisiológicos no logren presentárnoslo del modo en que lo hace el propio concepto fenoménico, tratado nuevamente como una clase de demostrativo.

No obstante, decíamos ya en el epígrafe que algo no funciona en la PCS. Las diversas caracterizaciones de los conceptos fenoménicos comparten el supuesto de que estos conceptos, a diferencia de cualquier otra clase de conceptos, no pueden definirse en términos de conceptos físicos ni su aplicación puede deducirse del conjunto completo de verdades acerca del mundo físico y verdades a priori. Esta asunción ha sido atacada por Tye (2009: 56 y ss.), que defiende que una enorme cantidad de conceptos comparten de hecho esa supuesta peculiaridad sin que a nadie, cabe añadir, parezca sugerirle este hecho la necesidad de abandonar una cosmovisión "materialista". Nosotros no desarrollaremos esta línea crítica porque, como indicábamos, nuestro rechazo de la PCS se basa en razones anteriores: la PCS acepta la jerga de la concebibilidad en la que los argumentos misterianos están formulados. Los motivos para desconfiar de la referencialidad de dicha jerga contarán así, al tiempo, como motivos para ahorrarnos la molestia de valorar la capacidad de cualquiera de las versiones de la PCS para hacer frente al desafío misteriano. Una vez desechamos la jerga de la concebibilidad, el desafío misteriano se revela apenas desafiante.

La PCS sirve para interceptar las conclusiones ontológicas que quepa extraer de los argumentos misterianos –que la naturaleza de la conciencia fenoménica y la del mundo físico son heterogéneas e irreconciliables, que pertenecen a esferas ontológicas diferentes–, pero no permite hacer nada con las conclusiones epistémicas que los filósofos misterianos pretenden alcanzar con sus argumentos –que la conciencia fenoménica no puede ser explicada científicamente–. De hecho, los defensores de la PCS no sólo aceptan que existe una brecha epistémica, sino que con su estrategia enfatizan las conclusiones epistémicas misterianas al imponer una infranqueable frontera conceptual entre lo físico y lo fenoménico.

Nuestra forma de hacer frente al desafío misteriano trae implícito el rechazo de la PCS: al desenmascarar la vacuidad de la jerga de la concebibilidad en la que dicho desafío se basa, la PCS se nos presentará como una crítica de la conclusión que acepta, sin embargo, las equívocas premisas de las que aquélla pretende extraerse. Expondremos a continuación los motivos por los cuales cabe desconfiar de la jerga de la concebibilidad incidiendo en aquéllos por los cuales la conexión entre lo concebible y lo posible que los misterianos postulan se muestra problemática.

Ya Parménides aventuraba esa idea según la cual "lo que cabe concebir y lo que cabe que sea son una misma cosa" (fragmento 3, en Bernabé Pajares, 2001: 156). No obstante, sus comentaristas contemporáneos apuntan que no debiera interpretarse este fragmento de forma ingenua como si el mismo tratara de expresar que el acto de pensar en una entidad o fenómeno cualquiera implicara la posibilidad de la existencia de dicha entidad o fenómeno, dado que el verbo *noein*, que debiera leerse en el sentido de percatarse de la verdadera naturaleza de una cosa, bloquea una tal interpretación (Ibíd.: 149). Con todo, el espíritu y la letra de una interpretación semejante es lo que hallamos en el espíritu y la letra de la jerga de la concebibilidad.

Los objetivos de los misterianos requieren que aquello que, según ellos, cabe concebir sea por ello posible. ¿Podemos asentir calmos a este supuesto? Descartes, en la meditación sexta, propone que "no puede dudarse que Dios tenga el poder de producir todas las cosas que [Descartes es] capaz de concebir con distinción", y añade que "nunca [ha] juzgado que le fuese imposible [a Dios] hacer cosa alguna sino porque [Descartes] encontraba contradicción en poderla concebir bien" (Descartes, 1641: *Meditaciones VI*). Persuasivo, pero no es muy difícil concebir el modo en que surgen con este tipo de planteamiento dos clases de problemas.

En primer lugar, qué sea concebible no es enteramente claro. Por ejemplo, ¿qué consideraremos contradictorio y por tanto inconcebible? ¿Sólo aquello que viole las leyes de la lógica? ¿Cuál lógica? ¿Por qué sólo esas leyes? Si admitimos que las leyes de, pongamos por caso, la física, tienen algo que decir acerca de lo concebible por lo que al movimiento respecta, ¿significaba «concebible» en tiempos de Descartes lo mismo que hoy día?[22] Partiendo del mismo supuesto, ¿qué crédito cabe conceder a las

[22] Usuarios de la jerga de la concebibilidad como Levine, curiosamente, aceptan la historicidad de nuestras intuiciones modales (Levine, 1998: 452), aunque, para guardarse las espaldas, se apresuren a distin-

conclusiones extraídas de argumentos que apelan a la concebibilidad en áreas científicamente incompletas? ¿Y cómo podría, por otra parte, decidirse cuándo cabe hablar de una determinada área de una disciplina científica como "completa"?

No disponemos de una definición consensuada de «concebible», como tampoco disponemos de métodos o criterios incontestables para decidir qué es concebible y qué es inconcebible. Parece que los usuarios de la jerga de la concebibilidad, al definir vagamente un escenario como concebible en tanto quepa imaginarlo sin incurrir en contradicciones lógicas, sólo pueden encontrar auspicio para su noción de lo concebible en la esfera de lo apriorístico y la analiticidad dinamitada por Quine (1951). Pero quizá nos quepa proponer ejemplos extremos, esto es, quizá todo el mundo estuviera de acuerdo en que un soltero casado es inconcebible. Sin embargo, ¿lo es asimismo el mayor número primo? ¿Y una mente inmaterial? ¿Y un cangrejo homeotermo? ¿Y uno mamífero? Parece claro que resulta más sencillo definir métodos y criterios para decidir si un animal es o no un mamífero que para decidir si un escenario es o no concebible: los casos límite en los que resulte imposible alcanzar un acuerdo acerca de si algo es o no concebible excederán con mucho los casos límite en los que resulte imposible alcanzar un acuerdo acerca de si un determinado animal cuenta o no como mamífero. La jerga de la concebibilidad supone, por otra parte, que podemos concebir todo aquello que la noción de réplica molécula a molécula de un ser humano implica, del mismo modo que supone que podemos concebir todos los fenómenos y teorías científicas relevantes para la explicación de la visión en color e, incuso, que podemos concebir a un ser humano capaz de manejar cabalmente semejante cantidad de información. De este modo, la jerga de la concebibilidad pretende hacer digerible la idea según la cual podemos imaginar a un ser humano que se sienta tranquilamente en su sofá y ante cuyo ojo de la mente aparece de golpe y porrazo, con total claridad y distinción, una perspicua maqueta mental compuesta por todos y cada uno de los átomos que integran un ejemplar de Asimo, todos y cada uno de los circuitos que integran su sistema de navegación, todos y cada uno de los electrones que circulan por éstos y todos y cada uno de los sucesivos marcos teóricos relevantes para la explicación del modo en que los mismos funcionan. Que seamos capaces de concebir tal cosa es algo un tanto incierto. Que podamos asegurar que alguien capaz de sentarse de ese modo en su sofá encontraría que la locomoción de Asimo es un misterio por cuanto después de aparecérsele ante el ojo de la mente el espectáculo com-

guir entre una suerte metafísica y otra epistemológica de posibilidad o, alternativamente, una suerte amplia y otra estrecha de concebibilidad.

pleto de la economía físico computacional del robot, acompañada de los sucesivos marcos teóricos en que la misma encajaría, no pudo encontrar ningún motivo para concluir que el robot caminaría es una mera –y curiosa– expresión de deseo. No sabiendo muy bien qué decimos cuando afirmamos que un escenario es concebible, menos podemos saber si lo que decimos cuando afirmamos que lo concebible es, *ipso facto*, posible tiene algún sentido. En esta situación, resulta difícil concebir el modo de extraer conclusiones sólidas de premisas enunciadas en jerga de la concebibilidad.

En segundo lugar, la segunda clase de problemas que entraña el uso de la jerga de la concebibilidad resulta más abstracta, y ello porque haremos para exponerla abstracción de lo ganado con lo antedicho. Concediendo que «concebible» tenga un sentido claro, ¿qué tipo de vínculo encontramos entre lo concebible y lo posible? Según los usuarios de la jerga de la concebibilidad, lo concebible es, *ipso facto*, posible.[23] No obstante, hemos podido comprobar desde principios del pasado siglo XX no ya la posibilidad sino la efectividad de hechos que atentan contra nuestras más arraigadas intuiciones de sentido común: a día de hoy no cabe dudar de la efectividad de fenómenos que resulta problemático afirmar que quepa concebir con distinción. Así, intuitivamente, por un punto externo a una recta sólo cabe trazar una paralela, pero dentro del marco de una geometría incapaz de deshacerse de esta intuición se hace imposible la descripción del universo actualmente aceptada en física.[24] ¿Qué haremos, pues, con el vínculo entre lo concebible y lo posible? ¿Dejar que funcione en dos sentidos, esto es, que valga tanto lo inconcebible pero posible –y, de hecho, efectivo: piense el lector en su dilecta perplejidad cuántica– como lo supuestamente posible *por* concebible? Entendemos que ésta no es una opción razonable y que, de hecho, tal y como Patricia S. Churchland ha subrayado (Churchland, 2002: 181), no contamos con ningún motivo para creer que lo concebible implique nada en absoluto acerca de lo posible en cualquier sentido relevante del término. "It seems to me that no serious argument can be based upon what we can or cannot imagine" (Russell, 1945: 715 del original, 770 de la traducción).

[23] Suele distinguirse entre lo metafísicamente posible y lo nomológicamente posible, siendo metafísicamente posibles todas las situaciones de las que podemos formarnos representaciones lógicamente consistentes y nomológicamente posibles las situaciones de las que podemos formarnos representaciones lógicamente consistentes que además sean acordes con las leyes que rigen el universo real. Por lo que a nuestra argumentación respecta, podemos prescindir de esta distinción. Señalemos que la misma no deja de resultar extraña cuando nadie sabe exactamente cuáles son las leyes que gobiernan nuestro universo.

[24] Al hilo de esta idea ha afirmado Remo Bodei que "al que entra en la ciencia se le pide que abandone la intuición" (Bodei, 1997: 35 de la traducción).

Más allá de los equívocos supuestos en los que encuentran sustento todos y cada uno de los argumentos misterianos, podemos aún preguntar qué marcha mal en cada uno de ellos.

3.2._Por qué no funcionan los argumentos misterianos

a) El argumento epistemológico y el argumento del punto de vista

El truco de manos, la falencia argumentativa común a ambos argumentos consiste en una tácita aunque nada inocente anfibología entre las nociones de experimentar y conocer destinada, además, a extraer, como Anselmo de Canterbury en el segundo capítulo de su *Proslogion*, conclusiones ontológicas (la falsedad del fisicalismo) de premisas epistemológicas. Por otra parte, a pesar de no hallarse explícitamente formulados en jerga de la concebibilidad, a ambos les es aplicable nuestra crítica de la misma en tanto el argumento de Jackson nos pide que imaginemos a alguien capaz de concebir todos los fenómenos relativos a la visión en color pero incapaz de figurarse la propia visión en color y en tanto el argumento Nagel hace uso de los dispositivos intuitivos en los que la jerga de la concebibilidad vendría con posterioridad a basarse al sugerir que todos los que él denomina "análisis reductivos de lo mental" serían lógicamente compatibles con la total ausencia de experiencia consciente, idea con la cual, como señalábamos, se anticipa al nacimiento de los zombis filosóficos –paradigma del uso y disfrute de la jerga de la concebibilidad– por cuanto a esta compatibilidad lógica ha de subyacer, necesariamente, una supuesta capacidad para concebir, por una parte, todo cuanto cualquiera de esos análisis reductivos pudiera contener y, por otra, a una criatura que, plegándose al mismo, careciera de experiencia consciente.

Respecto del aludido uso ambiguo de las nociones de conocimiento y experiencia, señalemos brevemente que la primera puede explicitarse mediante la referencia a conceptos, relaciones entre los mismos y actitudes de individuos tanto hacia unos como hacia otras: nos encontraríamos ante un marco, por así decir, reductible sin pérdida aparente a proposiciones (v. g.: «$4HCl + MnO_2 \leftrightarrow MnCl_2 + 2H_20 + Cl_2$», «Pedro sabe que $4HCl + MnO_2 \leftrightarrow MnCl_2 + 2H_20 + Cl_2$»). En el caso de la segunda noción, la posibilidad de reducir sin pérdidas una experiencia a proposiciones resulta más que discutible. Se trata de dos nociones ciertamente diversas y usarlas indiscriminadamente como si fueran intercambiables no puede sino conducir a equívocos. Cabe, por otra

parte, preguntar si esta diferencia entre conocer y experimentar implica la imposibilidad de explicar una experiencia haciendo uso de proposiciones y conceptos, a los cuales, apuntábamos, no parece poder reducirse, pero no debiera cundir el pánico: nada puede reducirse propiamente a proposiciones, salvo las proposiciones mismas. Una mesa es una mesa, pero –discúlpese la perogrullada– ni las proposiciones acerca de mesas ni el propio concepto de mesa pueden serlo.

Ambos argumentos intentan sacar parido de lo contenido en la última frase, esto es, parecen pretender probar que la experiencia consciente es inexplicable partiendo del escasamente interesante hecho de que una explicación de determinada experiencia consciente, aunque fuera correcta, no nos permitiría encarnarla. Pero se da el caso de que tampoco las más adecuadas entre las explicaciones de la caída del imperio romano o, pongamos por caso, de la rotación de Urano sobre un eje a escasos grados de su plano orbital nos permitirían encarnar semejantes fenómenos ni serían, en cualquier caso, equivalentes a los mismos. Breve y, en cierta medida, metafóricamente: de un *explanans* puede saltarse a un *explanandum*, pero nunca obtenerse un *factum*.[25] La explicación científica de la digestión ni puede digerir ni nos convierte en estómagos del mismo modo que la explicación científica de la experiencia consciente no nos permitirá tener esta o aquella experiencia consciente. Una explicación es una explicación, un fenómeno es un fenómeno. Decíamos que es un hecho que una explicación científica de una determinada experiencia consciente no nos permitiría en ningún caso encarnarla, y añadíamos que se trata de un hecho un tanto insulso. ¿Por qué? Pues porque a nadie se le escapa que los medios y objetivos de las ciencias y las artes difieren sustancialmente. Como Adam Zeman ha señalado a este respecto, las ciencias no tratan, como las artes, de conjurar de forma personal y específica la presencia viva de la experiencia, sino que, trabajando con principios y regularidades, intentan explicar lo que experimentamos, lo cual ni equivale ni pretende equivaler a evocarlo (Zeman, 2008: 197 del original, 186 de la traducción).

Ambos argumentos pretenden, por otra parte, probar la existencia de propiedades de los estados mentales que no pueden reducirse a hechos físicos de ninguna clase, con lo cual la generalísima tesis que denominan "fisicalismo" sería falsa. Esto, en el caso

[25] De ahí que quepa responder al misteriano, de acuerdo con el cual jamás podremos comprender la naturaleza de la conciencia, del mismo modo que Neils Bohr respondió a Werner Heisenberg, preocupado por la idea de que los seres humanos jamás pudieran comprender la naturaleza de los átomos: "Claro, todavía podemos lograrlo, pero para conseguirlo tendremos que aprender lo que significa en realidad el término «comprender»" (citado en von Weizsäcker, 1985/2006: 250τ), pues el mismo parece referir en el idiolecto misteriano indiscriminadamente, apuntando a nociones tan diversas como las de "explicar" o "encarnar".

del argumento epistemológico, se seguiría del hecho de que Mary conoce todos los hechos físicos acerca de la visión en color y, sin embargo, aprende algo nuevo al abandonar su cautiverio en blanco y negro y tener por vez primera experiencias cromáticas. La conclusión no se sigue, pues Mary estaría accediendo al mismo hecho (la visión en color) de dos modos distintos, conceptualmente en el primer caso y experiencialmente en el segundo, al abandonar su cautiverio acromático. La única conclusión plausible, aunque no particularmente interesante, es que leer un detallado libro sobre Praga no es como visitar Praga, lo cual no quiere decir que el libro haya de ser falso o insuficiente (podría tratarse de una colección en millones de volúmenes que contuvieran información precisa de hasta incluso el último rincón de la ciudad, de *todos* los detalles), sino que no va a transportar al lector a Praga más que metafóricamente. ¿Por qué? Por el mero hecho de que podemos tener un concepto acabado de Mercurio como primer planeta del Sistema Solar sin habernos acercado nunca a un telescopio y, a la inversa, podemos haber tenido experiencia sensible (visual) del planeta que los astrónomos llaman Mercurio sin saber de él nada en absoluto salvo que "veo un pequeño punto luminoso a través de este aparato". En otras palabras: tanto la experiencia como el conocimiento proposicional basado en conceptos apuntan a determinados aspectos del mundo, pero cada uno de ellos apunta de un modo distinto, proporcionando vías diferentes de acceso a los mismos fenómenos, de donde en ningún caso cabe colegir que con estas dos formas diversas de apuntar a unos y los mismos fenómenos nos hallemos ante bloques ontológicos estancos y discretos ni ante brechas epistémicas insalvables. La conclusión que podría extraerse del argumento de Jackson sería, cuando más, que ser informado proposicionalmente difiere de ser informado experiencialmente –o, cabría decir, en términos russellianos, conocer *por familiaridad*–, lo cual podría ser el caso incluso cuando en ambos casos se nos estuviera informando de lo mismo. Jackson, podría defenderse desde una concepción amplia de la referencia de «qualia», alcanza a mostrar intuitivamente que los *qualia* ligados a estados mentales con contenido proposicional difieren de los perceptivos, pero no que los *qualia* existan como aspectos no físicos de los estados mentales, a no ser que pretendamos que *nuestra* epistemología determine *la* ontología.

b) La brecha explicativa

Criticaremos brevemente a continuación los dos desarreglos conceptuales en que se basa la estrategia argumentativa destinada a sustentar la intuición de la existencia de una brecha explicativa entre el mundo fenomenológico y el físico. Subrayemos en primer lugar que la misma se alza sobre un confuso y ambiguo tratamiento de las nociones de explicación e identidad. Levine propone que nos sentimos perplejos cuando se nos informa de que X es idéntico a una experiencia consciente –siendo X la descripción correcta de la fisiología de esa experiencia– mientras no lo hacemos cuando se nos informa de que el agua es idéntica a $H_2O$. Su aserto, evidentemente, se limita a la consideración del modo en que "nos suenan" los enunciados de identidad entre términos del lenguaje cotidiano y conceptos científicos, cosa que nada parece decir acerca de la posibilidad de elaborar teorías científicas de las que derivar explicaciones. Las conclusiones que cabría extraer partiendo de una tal comparación del modo en que nos sentimos al ser expuestos a diferentes clases de enunciados de identidad podrían servir, acaso, para atacar posiciones ontológicas eliminativistas según las cuales un dolor de muelas no duele –esto es, no es nada más que un determinado estado neurofisiológico–, pero resulta difícil imaginar cómo podría con las mismas demostrarse el impasse epistémico al que cualquier intento de elaborar una determinada teoría científica se encuentra abocado. Levine pretende alcanzar conclusiones acerca de la posibilidad de elaborar una teoría científica de la que partir hacia una explicación basándose en la comparación entre la forma en que nos sentimos ante la posibilidad de situar en enunciados de identidad términos del lenguaje cotidiano referidos a fenómenos físicos junto a conceptos científicos, por una parte, y la forma en que nos sentimos ante la posibilidad de hacer lo mismo con términos del lenguaje cotidiano referidos a experiencias conscientes, por otra. A este fin, introduce subrepticiamente en sus consideraciones acerca de nuestras impresiones al ser expuestos a enunciados de identidad la noción de explicación sin dar ninguna explicación del modo en que la misma encaja en semejante marco. Lo único que hace al respecto es señalar que las identidades del tipo "el calor es idéntico a energía cinética" resultan explicativas mientras las identidades psicofísicas no, y lo único que cabe replicar es que una cosa es un enunciado de identidad o una reducción ontológica[26] por la cual redefinimos una palabra del lenguaje cotidiano en términos de fenómenos causales sub-

[26] Noción que no debe confundirse con la de *explicación reductiva*, una locución habitual en filosofía de la mente que alude a una forma de explicación causal en la cual una propiedad de alto nivel de un objeto queda explicada por el comportamiento de sus partes integrantes. Levine (vid., v. g., Levine, 2001: 84) introduce indiscriminadamente esta locución en contextos en los que discute acerca de enunciados de identidad. Sólo él podría decirnos si lo hace inadvertidamente o con la intención de producir efectos retóricos subliminales.

yacentes y otra diferente una explicación científica, bien la concibamos desde el modelo clásico de las leyes de cobertura, bien desde una aproximación causal como la de Wesley Salmon, bien en términos funcionales como los utilizados en las concepciones de la explicación elaboradas por filósofos de la biología o bien, incluso, desde un punto de vista pragmatista como el de Bas van Fraassen. Los enunciados de identidad que considera Levine implican reducciones ontológicas o meras equivalencias entre conceptos científicos y palabras del lenguaje cotidiano. Sin embargo, pretende partir de estas exiguas premisas hacia la fabulosa conclusión de que es imposible elaborar una determinada teoría científica de la cual derivar una determinada explicación utilizando como si fueran intercambiables nociones que tienen tan poco que ver como la de un enunciado de identidad y la de una explicación científica. Los segmentos de las teorías científicas que hemos de utilizar para derivar explicaciones son habitualmente –concédasenos el sardónico adverbio– cosas más complejas que «agua=$H_2O$». Si algún defensor de la estrategia argumentativa del *explanatory gap* lograra demostrar que cualquier explicación científica que cupiera formular partiendo de una teoría acerca de la conciencia debiera equivaler a una negación eliminativista de la misma, entonces cabría alguna posibilidad de que dicha estrategia sirviera para demostrar la imposibilidad de enunciar explicación alguna dentro del marco de cualquier teoría científica sobre la experiencia consciente. Mientras tanto, el injustificado modo en que dicha estrategia argumentativa se desliza de la noción de identidad a la de explicación debiera contar como argumento para desconfiar de su legitimidad.

¿Cómo se desliza, pues, la argumentación de Levine del comentario acerca del modo en que nos sentimos cuando somos expuestos a enunciados de identidad a la especulación acerca de lo explicable y lo inexplicable? Acabamos de indicar que Levine no aclara el modo en que pretende moverse de la noción de identidad a la de explicación, pero lo cierto es sí que ofrece unas escuetas observaciones al respecto cuando apunta que las identidades que no involucran conceptos referidos a eventos fenoménicamente conscientes son explicativas por cuanto las mismas hacen explícitos los roles causales asociados a los términos del lenguaje cotidiano implicados en el enunciado de identidad que sea el caso y por cuanto esos roles causales son todo lo que hay que entender al respecto. Según este planteamiento, la diferencia relevante (por lo que a la noción de explicación respecta) entre enunciados de identidad que implican términos del lenguaje cotidiano referidos a estados fenoménicamente conscientes y enunciados de identidad que no los implican estribaría en que mientras éstos se hallan integrados por términos

referidos a fenómenos cuya ontología se agota en roles causales, aquéllos tendrían en el lado fenoménico del enunciado de identidad términos que refieren a hechos cuya ontología excede el ámbito de aquello que cabe caracterizar apelando a roles causales. Desde el punto de vista del modo en que la estrategia argumentativa del *explanatory gap* se desliza así de los enunciados de identidad a la consideración de la posibilidad de derivar explicaciones de teorías científicas, habría que entender –a pesar del modo en que Levine elude las conclusiones ontológicas– que dicha estrategia argumentativa sirve para probar algo ciertamente evidente: que si la conciencia fenoménica fuera un epifenómeno, entonces no podría explicarse científicamente.

En segundo lugar, Levine y el resto de los autores que se hallan bajo el influjo de la intuición de la existencia de una brecha explicativa dan por sentado que una hipotética mente capaz de –a la luz de las leyes de la naturaleza que Dios dispuso antes de ponerse a descansar–[27] considerar todos los datos acerca de un determinado volumen de agua a 373º K y una atmósfera de presión no encontraría problema en concluir que ese volumen de agua hervirá –sino que, de hecho, entenderá que es necesario que hierva– mientras sí lo encontraría si se tratara de concluir, partiendo de un completo conocimiento de mi fisiología, si soy o no consciente. El usuario de esta estrategia argumentativa se compromete, pues, con una doctrina según la cual, aunque diéramos con la descripción y la teoría neurofisiológica adecuada para la explicación de la sensación subjetiva del dolor de muelas, nos encontraríamos perplejos al –llevar a efecto la descomunal tarea de– considerar dicha descripción y dicha teoría porque, a diferencia de lo que sucede con el resto de nuestro conocimiento científico, el *explanans* que con dicha teoría y dicha descripción pudiéramos elaborar no nos presentaría al *explanandum* del dolor de muelas como necesario, esto es, no encontraríamos necesario que los fenómenos a los que dicha descripción y dicha teoría aluden hayan de verse acompañados de experiencia consciente alguna. De este modo, la estrategia argumentativa del *explanatory gap* busca auspicio en una concepción aristotélica del conocimiento científico según la cual, cuando conocemos de forma científica un fenómeno o proceso, podemos estar seguros de su necesidad, de la imposibilidad de que tuviera lugar de otro modo (Aristóteles, *Segundos analíticos*, I, 71 b 9 & I, 74 b 9),[28] una concepción del conocimiento científico que, in-

[27] Lo cierto es que Levine suele desatender la necesidad de incluir el matiz que hemos introducido entre guiones: a menudo presenta su intuición de la brecha explicativa aludiendo, sencillamente, a la acumulación de datos físicos y desatendiendo la diferencia entre acumular datos e integrarlos en teorías.

[28] Según Jesús Mosterín, Aristóteles retorna con esta doctrina "a la concepción parmenídea y platónica de un saber riguroso y seguro acerca de un mundo necesario y eterno" (Mosterín, 2006: 201).

versamente, trae consigo encubierta la resurrección del dogma hegeliano según el cual "la verdadera explicación de todo fenómeno (…) equivale a la demostración de su necesidad lógica" (Berlin, 1939/1978: 59 del original, 72 de la traducción), de suerte que con ella, con esta concepción logocéntrica del conocimiento científico, nos encontramos a apenas un palmo del espíritu en que Pedro Abelardo leía en el evangelio de Juan "Al principio era el *Logos*".

Por una parte, ni esa mente omnipotente que la estrategia argumentativa del *explanatory gap* requiere es la nuestra ni sabemos qué podríamos concluir si lo fuera. Por otra, y concediendo que podemos concebirnos como siendo esa mente y gestionando todos los datos y marcos teóricos pertinentes –una concesión, sin duda, excesiva–, nada asegura que fuéramos a encontrar diferencias entre el caso de la evaporación y el de la conciencia, porque cabe de hecho imaginar que pudiendo manejar esa cantidad de datos y leyes científicas encontráramos ambos igualmente sorprendentes o predecibles. Nada asegura que las leyes de la naturaleza se presenten como necesarias a ojos de Dios. Tenemos, en definitiva, pocos medios para juzgar cómo se le presentarían las cosas a semejante mente omnipotente y ningún modo de probar que aquello que podemos explicar científicamente sea necesario, pero bastantes motivos para dudar de cualquier conclusión extraída de premisas de este tipo.

A Levine y sus secuaces les parece que algo permanecerá inexplicado sea cual sea la cantidad de datos y marcos teóricos empleados para dar cuenta del modo en que la neurobiología deviene fenomenología. Algo les sorprende en ese paisaje causal al punto que consideran que será por siempre ininteligible. A nosotros nos sorprende lo trivial, inteligible y auto-transparente que les resulta cualquier otro paisaje causal, porque de hecho, en el fondo, no tenemos ni idea de por qué ni de cómo –si como necesarias o como contingentes– existen las regularidades naturales que se hallan a la base de ambas clases de paisaje. Lo único que sabemos es que esas regularidades tienen lugar y lo único que nos es dable investigar es el modo en que tienen lugar y el modo en que se relacionan entre sí: en eso consiste la ciencia. Desde luego, lo que ciertamente desconocemos es si esas regularidades son o no necesarias. Quizá Dios lo sepa. Por nuestra parte, sencillamente, podemos tener una u otra impresión al ser expuestos a nuestras conceptualizaciones de las mismas y decirnos algo así como "en vista este *explanans* me parece como si el *explanandum* se siguiera (o no) necesariamente", pero este tipo de afirmaciones poco dice acerca de la naturaleza de los fenómenos a los que *explanans* y *explanandum* se referirían o acerca de la posibilidad de elaborar teorías explicativas para los

mismos, sino que parecen decir más bien algo acerca del modo en que determinados animales enculturados en determinados medios responden a determinados conjuntos de enunciados. Muchos filósofos de la mente han contribuido a elevar especulativos edificios teóricos –como la PCS– tratando de dar cuenta de los motivos por los cuales la exposición a argumentos misterianos como el de la brecha explicativa conduce, de forma en apariencia ineludible, a una cierta sensación de incredulidad o desconcierto que cabría expresar mediante un "¡la conciencia parece, efectivamente, irreconciliable con el mundo físico!", pero han elevado esos edificios desde la base de la pertinencia de esa sensación, es decir, asumiendo que la misma obedece a –y se halla de hecho con recta justedad justificada por– algún fenómeno realmente existente y relacionado, por ejemplo, con el modo en que usamos conceptos o accedemos a datos referidos a propiedades físicas o propiedades fenoménicas. En lugar de seguir añadiendo pisos a esos edificios, entendemos que resultaría más interesante investigar –como recientemente ha hecho Paul Bloom desde la psicología del desarrollo (Bloom, 2004)– acerca de los motivos por los cuales tendemos a encontrar irresistiblemente atractiva la concepción dualista del ser humano que, de un modo u otro, sustenta tanto la formulación de los argumentos misterianos como esa extendida recepción de los mismos mediada por esa cierta sensación de desconcierto.

c) La clausura cognitiva

Incluso a Thomas Nagel se le antoja excesivo el diagnóstico de McGinn. "Aunque pudiera estar en lo cierto –dice–, creo que su pesimismo es prematuro" (Nagel, 1993a: 40τ). El pesimismo de McGinn parece apoyarse en una argumentación, pero lo cierto es que su argumentación sólo le sirve para ahondar en su intuición de que el problema de la conciencia es irresoluble y en su sensación de perplejidad ante el hecho de que algo como la experiencia consciente sea ocasionado por el funcionamiento de algo como el cerebro, una sensación de perplejidad que constituye el basamento de su postura misteriana y que McGinn (2002) presenta como una suerte de epifanía espiritual (Ross, 2008: 113). De hecho, y contra lo que muchos pudieran pensar al leer sus textos de los últimos once años del pasado siglo XX, McGinn no cree haber probado que el problema de la conciencia sea científicamente inabordable (vid., particularmente, la introducción a McGinn, 2004), sino que, entiende, se ha dedicado, sencillamente, a profundizar en

esa intuición y esa perplejidad mediante remedos de argumentos.[29] Con estos remedos de argumento no se propone demostrar de forma inconcusa la validez de su planteamiento sino, cuando más, dar pábulo a la señalada intuición en el corazón de aquellos lectores que, de algún modo, la hubieran acariciado ya. No tratándose, pues, de argumentos en los que las conclusiones pretenden seguirse de las premisas, la intuición de McGinn puede seguir en pie a pesar de lo que quiera que uno diga acerca de sus textos y los remedos de argumento que ellos contienen. De este modo, no cabe refutar la intuición de McGinn atacando sus "argumentos". Todo lo que cabe hacer es, como Nagel, dar mayor o menor crédito a sus vaticinios.[30] No obstante, McGinn apoya su intuición en una serie de ideas que sin duda resulta interesante valorar.

En primer lugar, el pronóstico de McGinn respecto de la imposibilidad de ofrecer respuestas sólidas a los problemas de los que se ocupa la filosofía no resiste un cabal cotejo con nuestra historia intelectual. Así, la investigación en física o biología viene iluminando problemas que hace escasamente un siglo eran considerados propiamente filosóficos e inaccesibles a las disciplinas particulares, como el origen del universo, la naturaleza del tiempo o la de la vida. McGinn plantea en este sentido que entre los asuntos con los que los filósofos pierden el tiempo y las cuestiones que cabe abordar científicamente hay un hiato insalvable, pero, ciertamente, nadie puede adelantar en qué áreas topará nuestro inquirir con hiatos entre lo que puede llevarse al laboratorio y lo que nunca podrá y, paralelamente, parece que hará falta multiplicar por varios órdenes de magnitud la elocuencia de Demóstenes para soñar con atisbos de consenso acerca del lugar exacto en que trazar la frontera entre las cuestiones propiamente filosóficas y las propiamente científicas, e incluso acerca de la pertinencia de dicho trazado. En cualquier caso, cuando se trata de especular –y, téngase presente, toda especulación puede tratar de ocultarse bajo el disfraz de diversos artefactos retóricos–, bien puede hacerse en un sentido u otro, del mismo modo que cuando se trata de intuir cada cual tendrá sus intuiciones. Así, nada favorece la intuición de que aquello que otrora fuera concebido como un inabordable arcano filosófico –pongamos, el problema de la libertad– mañana encajará perfectamente en un marco teórico sólidamente asentado en evidencias empíricas frente a la intuición de que, por el contrario, tal cosa jamás

[29] McGinn (2012) llega a sugerir que hacer filosofía de la mente consiste, precisamente, en permitirse sentir esta perplejidad.

[30] Mientras McGinn tendría la honradez de suscribir lo contenido en este párrafo, sería quizá el único misteriano en hacerlo –a pesar de que quepa decir exactamente lo mismo del resto de los argumentos misterianos.

sucederá. Le cabe al filósofo que se enfrenta hoy a estos problemas fronterizos contribuir a la investigación de los mismos o alimentar el brillo de la mística aureola de los inextricables enigmas de su predilección. Habrá de ser su intuición la que le indique cuál de estos caminos escoger y, nuevamente, poco o nada puede argumentarse a favor o en contra de mi intuición según la cual el sándwich Elvis era el desayuno favorito de la población de la Atlántida.

En segundo lugar, el primer puntal intuitivo con que McGinn pretende reforzar su corazonada de que el problema de la conciencia es, de hecho, un misterio es la aparente desconexión entre los fenómenos a los que accedemos introspectivamente y los fenómenos que revela la investigación neurofisiológica. Ambas clases de fenómenos guardan un mutuo silencio, propone McGinn. Igualmente, el magnetismo y la electricidad parecían intuitivamente dos clases de fenómenos enteramente distintos hasta los experimentos de Ørsted y Faraday y las ecuaciones de Maxwell. Ambas clases de fenómenos guardaban asimismo un circunspecto silencio mutuo. Fue necesario experimentar y elaborar marcos teóricos para que dejaran de hacerlo. Nada en la argumentación de McGinn impide suponer que algo análogo haya de suceder con el avance de la investigación científica de la conciencia –y los hay que, con los Churchland, afirman que, de hecho, viene ya sucediendo.

En tercer lugar, respecto del argumento de McGinn según el cual el conocimiento de la propiedad natural del cerebro que da lugar a la experiencia consciente habría de conferir también una comprensión de la teoría que deba dar cuenta del modo en que esa propiedad traza el vínculo entre fisiología y fenomenología, McGinn habla de dicha propiedad en dos sentidos incompatibles, como una propiedad general del cerebro que se hallaría a la base de nuestra experiencia consciente, por una parte, y como estados específicos del sistema nervioso que se hallarían asimismo a la base de experiencias conscientes específicas, por otra. Este equívoco le permite proponer que para entender la teoría que explica cómo la conciencia depende de aquella propiedad tendríamos que poder entender primero los términos en los que la misma se encuentre formulada, y dado que uno de ellos sería, según McGinn, el de la cualidad subjetiva de un estado cerebral –es aquí donde la propiedad general de cerebro se convierte en un estado específico de nuestro sistema nervioso–, esto es, un concepto del que carecemos dados los límites fenoménicos de nuestra capacidad de formar conceptos, conocer dicha propiedad nos conferiría un conocimiento que no podemos tener.

Con independencia de la ambigüedad terminológica que permite a McGinn desarrollar sus intuiciones, no es enteramente claro cómo pretende poner éstas al servicio de sus conclusiones misterianas. En este sentido, ¿por qué motivos conocer la propiedad a la que alude McGinn implicaría comprender la teoría en que la misma pudiera articularse para dar cuenta del modo en que la actividad neurofisiológica deviene conciencia fenoménica? McGinn no lo explica y, de hecho, antes de explicar esto debiera aclarar qué entiende por propiedad y qué entiende por comprender –o, en sus términos, captar (grasp)– una propiedad. Al no aclarar a qué se refiere con estas nociones, la más amable entre las interpretaciones que cabe hacer de su aserto sería la siguiente: al comprender *exhaustivamente* que algunos materiales conducen la electricidad (tienen esa propiedad) uno debiera estar en situación de formular las ecuaciones de Maxwell sin tener noticia previa de ellas. Desde el punto de vista de esta interpretación, cabe imaginar que comprender *exhaustivamente* cualquier propiedad implique –en el sentido de traer consigo–, de algún modo, conocimiento de los sucesivos marcos teóricos en que la misma debiera integrarse, pero no habiéndonos aclarado McGinn que hemos de entender por "comprender una propiedad", "comprender *exhaustivamente* una propiedad" resulta algo aun más oscuro. Pero incluso dando por sentadas ambas nociones, persiste un problema: sencillamente hemos *imaginado* que comprender una propiedad implica una suerte de iluminación agustiniana respecto de una serie de teorías científicas. Con todo, asumiendo la pertinencia del señalado acto de imaginación, precisar el sentido del adverbio «exhaustivamente» en el contexto en que venimos utilizándolo es imposible: quizá mañana una sorprendente teoría de la gravedad cuántica modifique nuestro concepto de una propiedad que, considerábamos, habíamos llegado a comprender exhaustivamente.

Yendo más allá, la más amable interpretación de lo que significa comprender una propiedad que se nos ocurre podría expresarse como sigue: comprender una propiedad significa tener conocimiento de todas las potencialidades causales asociadas a la misma. Pero conocer las potencialidades causales asociadas a una propiedad no parece ni ser lo mismo que conocer la teoría en la que las mismas se hallen articuladas y el modo en que ésta encaja en sus sucesivos marcos teóricos y se relaciona con otras teorías, ni implicar dicho conocimiento: da la impresión de que una cosa es saber cómo se comportará un sistema y otra distinta hallarse en disposición de explicar por qué se comportará de ese modo, del mismo modo que da la impresión de que aquel conocimiento no implica éste. Ilustremos este punto con un ejemplo. Tanto Dmitry Ivanovich Mendeleyev como Julius Lothar Meyer –el Russel Wallace de la tabla periódica, por así decir– habían capta-

do (grasp) perfectamente una importante propiedad de los elementos químicos: la periodicidad de sus características en función de sus pesos atómicos. Tanto el uno como el otro habían relacionado correctamente dicha propiedad con las distintas potencialidades causales de los elementos dentro de cada grupo, pero ni el uno ni el otro supieron explicar a qué se debía esa periodicidad. Conocían perfectamente el qué, esto es, la naturaleza periódica de los elementos químicos, pero esto no les sirvió para inferir el por qué de esa naturaleza periódica, y ello debido a que, sencillamente, no buscaron en el lugar apropiado al considerar que el peso atómico debía ser el factor decisivo en la elaboración de una teoría explicativa de la señalada periodicidad.

El resto del argumento, esto es, esa intuición de McGinn según la cual para entender la teoría que explica cómo la conciencia depende de aquella propiedad del cerebro –que ora parece plantear McGinn como propiedad general ora como estado particular– tendríamos que entender primero la noción de "cualidad subjetiva de un estado cerebral", confesémoslo, no lo comprendemos. Según McGinn, esta noción debería encontrarse entre los términos en los que la teoría capaz de dar cuenta del modo en que la conciencia depende de la señalada propiedad se encuentre formulada, pero en ningún momento explica por qué debiera hacerlo. En cualquier caso, parece que McGinn, a pesar de tratar de persuadirnos de que una tal teoría nunca nos será accesible, ¡resulta que tiene información privilegiada acerca de los términos en los que la misma debiera formularse!

En cuarto y último lugar, McGinn culmina su argumentación misteriana, como señalábamos, sumándose a los usuarios de la jerga de la concebibilidad cuando propone que la conciencia no juega ningún papel en la explicación de los hechos neurofisiológicos y no es, por tanto, necesaria para dar cuenta de los mismos. Como sugeríamos, McGinn se suma con esta afirmación a los usuarios de la jerga de la concebibilidad por cuanto la misma supone que el filósofo británico ha logrado concebir un hipotéticamente completo conjunto de fenómenos y modelos explicativos neurobiológicos y que, además, ese acto de concebir le ha servido para advertir que en dicho conjunto la conciencia no juega ningún papel. Nuestros motivos para dudar de las asunciones que subyacen al empleo de la jerga de la concebibilidad siguen vigentes ante este intuitivo dispositivo retórico. Pero, incluso haciendo abstracción de tales motivos, dicho dispositivo sigue mostrándose problemático. ¿Por qué? Sencillamente por el hecho de que McGinn obvia que la de conciencia es una noción ineludible en neurociencias. La misma desempeña un papel central en la elaboración de hipótesis, el diseño de experimentos y la ex-

plicación y discusión de los resultados obtenidos en neurociencia cognitiva de la memoria[31] (vid. Eichenbaum, 2002/2012; Parker et al., 2002), la atención[32] (vid. Posner, 2004/2012) e, igualmente, en neurociencia afectiva (vid. Armony & Vuilleumier, 2013; Lane & Nadel, 2002; Panksepp, 1998). La centralidad y necesidad de la noción de conciencia en neurociencias la evidencian, por otra parte, las más de 300 ocasiones en las que las voces «conscious» o «consciousness» son utilizadas en la tercera edición del manual *Fundamental Neuroscience* (Squire et al., 2008), las asimismo más de 300 en las que las mismas aparecen en la cuarta edición de *Principles of Neural Science* (Kandel, et al., 2000) o las más de mil en las que pueden leerse en la segunda edición de *The New Cognitive Neurosciences* (Gazzaniga, 2000). Pongamos dos escuetos ejemplos. Todo el mundo conoce de primera mano el grado en que difieren las experiencias de aprender una secuencia motora compleja (como tocar una pieza al piano) y ejecutar esa secuencia de forma automática una vez aprendida. Pues bien, es imposible diseñar un experimento de neuroimagen destinado a evaluar las diferencias en la activación de diversas áreas encefálicas durante las fases de entrenamiento y ejecución automática sin hacer uso de la noción de conciencia del mismo modo que es imposible discutir o explicar los resultados obtenidos en el mismo sin utilizar dicha noción. Análogamente, parece imposible realizar estudios neurofisiológicos sobre rivalidad binocular (vid., v. g., Blake & Logothetis, 2002) sin usar la noción de conciencia. Sin los datos que la psicología cognitiva les ofrecen, las neurociencias carecerían, entre otras cosas, de *explanandum*, y a nadie se le escapa que es difícil dar tres pasos seguidos en psicología cognitiva sin topar con la conciencia. En esta situación, a muchos nos resulta difícil concebir un hipotéticamente completo conjunto de fenómenos y modelos explicativos neurocientíficos en los que la conciencia no aparezca por ninguna parte, porque la misma aparece ya constantemente en el actual conjunto incompleto de los mismos. Puede que la total desatención de McGinn al constante refinamiento de los métodos y marcos teóricos en las interdisciplinas a caballo entre la psicología y la biología (como la psicobiología, la neuropsicología, la neurociencia cognitiva y la afectiva, la psicofisiología o la psicofarmacología) se halle a la base de su capacidad para imaginar del señalado modo ese hipotéticamente completo conjunto de fenómenos y teorías, y quizá sea, en último término, esa total desatención el aspecto más censurable del modo en que McGinn preten-

[31] Particularmente por lo que respecta a la distinción entre formas implícitas y explícitas de memoria y los estudios de neuroimagen que explotan los diversos paradigmas de priming.
[32] Nuevamente la noción es aquí indispensable en la elaboración y discusión de estudios de priming (vid. Albright et al., 2000: 46-47).

de con sus intuiciones haber sondeado las limitaciones con que, por principio, toparán esas interdisciplinas.

d) Los zombis y los espectros invertidos

Estos dos últimos grupos de argumentos son netamente dependientes de la jerga de la concebibilidad. Rechazada ella, rechazados ellos. Como señaláramos, estos argumentos se basan en una injustificada intuición según la cual que los grotescos escenarios que presentan sean concebibles implica que son por ello posibles. Como hicimos notar, la noción de lo concebible tal y como circula por estos círculos es, por decir lo menos, problemática. Además, no encontramos fundamento alguno para la suposición de que lo concebible implica algo acerca de lo posible en algún sentido relevante. Por otra parte, ambas clases de argumento perderían sustento y objeto toda vez que nuestra argumentación acerca de la ilegitimidad de la propiedad de la intrinsecalidad como característica de nuestra noción de *qualia* alcance su blanco. Estamos, pues, enteramente de acuerdo con Levine (1995) cuando señala que abandonar la intrinsecalidad como rasgo de nuestra noción de *qualia* conlleva abandonar también nuestras intuiciones acerca de la posibilidad de espectros invertidos y *qualia* ausentes. No obstante, a pesar de entender que lo señalado es suficiente para depositar en esta clase de argumentos algunas de nuestras más razonables dudadas, fijémonos brevemente en cada uno de ellos.

Respecto de los argumentos basados en la concebibilidad de zombis o *qualia* ausentes, apuntemos brevemente que es ciertamente dudoso que alguien alcance a concebir *todo* lo que implica la noción de "réplica molécula a molécula de un ser humano", porque es, de hecho, dudoso que alguien alcance a concebir acabadamente una molécula, entre otras cosas porque aún seguimos preguntándonos cómo hemos de concebir un átomo.[33] Por otra parte, recuérdese que los argumentos basados en la concebibilidad de zombis apelan a una hipótesis según la cual aun cuando fuéramos capaces de concebir todo hecho relevante en relación con la base física y funcional de un zombi réplica exacta de un ser humano consciente, no tendríamos motivos para

[33] El supuesto de la concebibilidad de los zombis ha sido criticado de forma tan extensa en la literatura que no consideramos necesario insistir en él. Algunas de las más elocuentes entre estas críticas han sido formuladas, precisamente, por el introductor del término, seducido en los setenta por la corazonada zombi y sus implicaciones antimaterialistas y, actualmente, uno de los más acérrimos detractores de la misma (vid., v. g., Kirk, 1994; 1999; 2005).

atribuirle conciencia fenoménica. Incidamos en que es más que dudoso que nadie pueda concebir tal cosa, pero, dejando al margen estas dudas, preguntémonos por los motivos que pudieran hallarse a la base de la aseveración del que dice haber llevado a cabo semejante proeza sin hallar motivos para atribuir conciencia a dicho ser imaginario. ¿Puede alguien desprovisto de la intención de resucitar la hipótesis dualista conceder que *todos* los hechos físicos, funcionales y relacionales relativos a cualquier fenómeno natural son menos de lo que tal fenómeno natural es en sí mismo? Resulta ciertamente esclarecedor plantear esta pregunta desde la perspectiva que ofrece la noción de "zagnet" (Dennett, 1993b: 211), un "imán zombi", idéntico en todos los sentidos a un verdadero imán, un "imán" que se comporta por tanto como un verdadero imán, pero un imán entre comillas, en definitiva, por cuanto carece de una misteriosa cualidad interna, una esencia oculta –el alma que Tales atribuía a los imanes, digamos–. Nadie sabe, en definitiva, cuál sería el resultado de un acto de imaginación como el que piden poner en juego los argumentos basados en zombis (consistente, recordemos, en nada menos que "concebir" hasta el último detalle físico de un ser humano), pero, desde luego, nada asegura que esos resultados fueran a serle favorables al misteriano, y de hecho hay quienes opinan que, con el avance de las diversas ramas de las neurociencias, es cada día más difícil concebir un universo idéntico al nuestro y poblado por criaturas idénticas a nosotros pero incapaces de ser sujetos de experiencia consciente (Churchland & Churchland, 1997), siendo así que, a medida que profundiza uno en las disciplinas pertinentes, le resulta cada vez más complicado no ya dar crédito, sino incluso comprender esta clase de argumento misteriano.

Por su parte, lo que los argumentos basados en espectros invertidos nos piden que imaginemos es más de lo que a simple vista parece. Al presentarlos indicamos que lo que nos piden que imaginemos es que individuos con experiencias diversas –invertidas– reaccionan íntegramente de forma idéntica ante los mismos estímulos, y añadimos que esto no se limitaría a la conducta verbal, sino que atañería hasta a la última de las disposiciones reactivas implicadas en el procesamiento de la información relativa a un estímulo (incluso índices de respuestas emocionales implícitas como la respuesta galvánica cutánea habrían de mantenerse constantes mientras varían sólo los *qualia* para que el argumento alcanzara sus objetivos). Lo que nos piden, pues, es que imaginemos una serie de propiedades de los estados mentales netamente independientes de todo cuanto las rodee o constituya –otro tanto cabría decir de lo que los argumentos basados en la concebibilidad de *qualia* ausentes nos piden que imaginemos–. Rechazada la

propiedad de la intrinsecalidad de los *qualia* ante las dudas razonables suscitadas por la idea de algo aislado y causalmente inerte, la misma suerte habrá de correr la idea del espectro invertido.

Destaquemos para terminar que existen motivos empíricos que bloquean la posibilidad de espectros invertidos. Puede que los mismos sean posibles en algún extraño sentido y en algún mundo posible, y puede que esa posibilidad implique algo respecto de ese mundo, pero nada de eso parece suceder en el mundo que nos concierne, esto es, el de la fenomenología y la fisiología humana. Como el resto de los argumentos discutidos hasta aquí, el del espectro invertido trata de decidir a priori con cuántas posibilidades de éxito contarían diferentes programas de investigación empírica basándose en intuiciones discutibles y desprovistas de soporte empírico. En su caso, partir de esas discutibles intuiciones acarrea un serio problema, porque el ámbito empírico al que trata de encasquetar dichas intuiciones no se pliega a las mismas. Nuestro espacio cromático no es reversible en la forma lineal y simplista que el proponente del argumento del espectro invertido presume dado que, en último término y en pocas palabras, existe una diferente cantidad de cada una de las tres clases de células sensibles a la longitud de onda que pueden encontrarse en nuestra retina y, adicionalmente, un específico solapamiento entre las longitudes de onda a las que cada una de esas tres clases de célula responden, lo cual resulta en una mayor capacidad para discriminar determinados sectores del espacio cromático, lo cual, a su vez, impide una inversión del espacio cromático fenomenológico, por así decir, *de uno a uno* que no dejara rastros conductuales o reactivos, esto es, una inversión que permitiera al sujeto de la misma superar sutiles pruebas discriminatorias sin desconcertar profundamente a los expertos que las administraran y examinaran los resultados de las mismas (vid. Churchland, 2002: 183 y ss.).

La conclusión que cabe extraer de nuestro recorrido a través de la tentativa misteriana de convencernos del lugar al que, en cualquier caso, les cabrá llegar a nuestras explicaciones es que nadie tiene la menor idea de cuál es ese lugar. Ciertamente, es más que probable que la potencia explicativa de las ciencias y, con ella, nuestro acceso epistémico a la realidad, vayan a ser por siempre limitados, pero nadie conoce sus límites.

*Habiendo concluido que no existen motivos para el desánimo y que el de explicar científicamente la conciencia es un proyecto en principio hacedero, pasaremos a argumentar acerca de la posibilidad de dicha explicación y acerca de aquellas tendencias, inercias, propuestas y disputas teóricas que tienden a favorecerla y aquéllas que tienden a entorpecerla. El primer capítulo de la subsiguiente segunda parte proporciona el contexto para dicha argumentación. En él exponemos y defendemos la postura naturalista que subyace a la misma. El segundo, por su parte, lo dedicamos a denunciar las apuntadas tendencias teóricas y los enredos conceptuales que vienen entorpeciendo el avance hacia la comprensión de la naturaleza de la conciencia, mientras que en el tercero trataremos de arrojar luz sobre las vías que cabe esperar que conduzcan en la dirección opuesta habilitando la posibilidad de dicha comprensión.*

# II

# PARTE SEGUNDA

## Replanteamiento del núcleo del debate contemporáneo en torno al problema de la conciencia

# CAPÍTULO 7

## EL NATURALISMO COMO PIEDRA DE TOQUE DE NUESTRO REPLANTEAMIENTO DEL NÚCLEO DEL DEBATE CONTEMPORÁNEO

### 1._Filosofía naturalista

Dedicamos este capítulo a definir la noción de naturalismo y defender la postura naturalista a la que nos adherimos para aplicarla en el siguiente al problema de la naturalización de la conciencia. La tarea de delinear un marco histórico y conceptual capaz de facilitar una cabal comprensión de eso que debiéramos entender por naturalismo en filosofía podría obligarnos a postergar el tema del que hemos de ocuparnos indefinidamente, invitándonos a excursiones históricas y excursos epistemológicos que, de ser abordados con la mínima escrupulosidad, afectarían negativa e irrefragablemente a la unidad, la extensión y los propósitos de esta tesis. No obstante, esbozar de forma compendiosa dicho marco será de utilidad de cara a aproximarnos al substrato de nuestro replanteamiento del debate contemporáneo en torno al problema de la conciencia.

### 1.1._ Sucinto bosquejo de la génesis de la filosofía naturalista

El naturalismo es, ante todo, una actitud: la del que comprende las signaturas en los anaqueles, las facultades en las universidades y los departamentos en las facultades no como una compartimentación de la actividad investigadora e intelectual que vendría a reflejar otra análoga en el mundo real, sino como resultado de una cierta economía léxica y unas determinadas condiciones socioeconómicas. Una actitud, pues, y también una perspectiva acerca del significado del término «filosofía». Según esta perspectiva,

no existe una diferencia esencial entre el trabajo de un lógico y el de un biólogo. Ambos emplean su tiempo en quehaceres filosóficos. Esta actitud y esta perspectiva pueden entenderse también como una suerte de retorno, como un camino de vuelta hacia la concepción de la actividad investigadora e intelectual que acompañó a dicha actividad desde sus orígenes. En este sentido, Aristóteles hubiera dicho de sus estudios lógicos y sus estudios biológicos que ambos fueron resultado de un mismo ejercicio: el del pensamiento contemplativo. Comenzaremos a trazar una brevísima historia de este retorno en el punto en que tradicionalmente vino presumiéndose que se produce una crucial ruptura en sentido contrario y las ciencias se "independizan" de la filosofía: en la época de la revolución científica. Recorreremos a grandes pasos esa historia con un ojo sobre sus contornos generales y otro sobre los del estudio de lo mental.

Con el periclitar de la hegemonía escolástica tuvo lugar una imponente eclosión intelectual caracterizada por una nada desdeñable –contra lo habitualmente indicado en la historia oficial, la historia del erizo (Berlin, 1951)– diversidad de propuestas metodológicas y teóricas –empirismo, mecanicismo e inducción en Bacon, deducción en Descartes y resolución/composición en éste y Galileo–, todas ellas catalogadas por la manualística historiográfica bajo una misma etiqueta: "época de la revolución científica". Efectivamente, durante el siglo XVII se produjo un cambio de dirección en el proceder de los *filósofos naturales* que, según las reconstrucciones históricas contemporáneas, fue determinante de cara a sentar las bases para el desarrollo de la ciencia moderna. Asimismo esta época cimienta una moderna interpretación materialista y mecanicista de aspectos de la experiencia humana que tradicionalmente se concibieran como inaccesibles al análisis científico y que comienzan ahora, empero, a contemplarse como susceptibles de explicación dentro del marco de las emergentes ciencias naturales. Hobbes –cuya concepción materialista del hombre hallaría continuidad en el siglo siguiente de la mano de La Mettrie o d´Holbach– es, por su radical planteamiento de lo mental en términos mecanicistas, tal vez el ejemplo más explícito de este tipo de concepciones en esta época.

Al periodo en el que la actividad, la investigación y la experimentación científica dieron el giro al que nos hemos referido como un cambio de dirección suele arrogársele la mencionada problemática signatura: "época de la revolución científica". En lo atinente a lo que aquí nos concierne, cabe señalar que en esta época puede igualmente encontrar el historiador de la filosofía el germen de una filosofía naturalizada, fuertemente influida por teorías y resultados científicos contemporáneos. Pero, por lo que a éstos

toca, esto es, en lo relativo a las teorías, los resultados y, particularmente, la metodología de la ciencia, ¿qué han encontrado en esta época los más recientes estudios de historia de la ciencia? Puede que la respuesta a esta pregunta arroje alguna luz sobre la pertinencia del modo en que tanto los filósofos que hoy se proclaman naturalistas como sus detractores se refieren a esta época seminal. Tratemos de ofrecer la respuesta corta a la misma antes de acercarnos, en orden cronológico, a programas naturalistas más recientes.

Sin tratar de negar o restar significación al alcance de este periodo crucial para la génesis de las ciencias y la epistemología modernas, algunos intérpretes, como el historiador de la ciencia Steven Shapin,[1] han tratado de matizar y atemperar el tal vez excesivo énfasis en que habitualmente se incide al hablar de este periodo en términos de drástico y unívoco punto de inflexión en la historia de la ciencia. Lo que dichos intérpretes apuntan, resaltando junto a lo innovador lo conservador de las ideas de los autores de esta época, es que este periodo no supuso una ruptura tan radical con el aristotelismo medieval como sus propios protagonistas pretendieron. A pesar del afán trasgresor e innovador de aquellos *filósofos naturales* que en el siglo XVII se propusieron comenzar una reconstrucción total del edificio de las ciencias, no puede afirmarse que existiera entonces unidad de opinión respecto del carácter novedoso de los emergentes criterios metodológicos ideados como arietes con los que abrir las puertas del estudio riguroso de la naturaleza. Sí puede, por el contrario, hallarse en estos *filósofos* del XVII una relativa uniformidad en otro respecto: la pretensión de erigir un nuevo cuerpo de saber científico, una nueva forma de abordar el estudio de la naturaleza, pretensión a la que respondiera su porfiada insistencia en la importancia de la experiencia frente al tradicional peso de la autoridad y el testimonio. Comenzaron, pues, aquellos *filósofos* a dar prioridad al libro de la naturaleza, anterior y más rico que el escrito por cualquier mortal. En él trataron de leer mediante la experiencia, la observación directa y –en conformidad con el célebre adagio galileano– la matematización.[2]

Como puede apreciarse, en este momento incluso la nomenclatura parece desdibujar las fronteras entre ciencia y filosofía. Así, por ejemplo, el cinco de julio de 1687 aparece publicada la obra que inaugura la mecánica clásica bajo el título *Philosophiæ Naturalis Principia Mathematica*. Quince años antes, Isaac Newton, autor de la misma, enviaba sus primeros artículos sobre óptica a la primera revista científica de la historia,

[1] Para el extremo que aquí tocaremos vid. Shapin (1994; 1996).
[2] En lo que a este punto se refiere, vid. Montesinos Sierra (2004).

que ha conservado su nombre original hasta la actualidad: *Philosophical Transactions of the Royal Society*. Esta falta de claridad en el trazado de demarcaciones sigue observándose en siglos posteriores. En 1808 John Dalton publica *A New System of Chemical Philosophy*, la primera fundamentación atomista de la química; la cátedra que James Clerk Maxwell ocupara en el Marischal College de Aberdeen entre 1856 y 1860 era la de "Natural philosophy"; todavía hoy en los países de habla inglesa el título de Doctor en cualquier disciplina científica se denomina PhD (Philosophiæ doctor).

Los *filósofos naturales* del siglo XVII consideraron, en cualquier caso, que para la consecución de un saber científico sobre el mundo natural, un saber cierto, se hace necesaria una forma adecuada de acceder al mismo. La experiencia, aunque se entendía que debía ser incorporada a los fundamentos del conocimiento científico adecuado, sin más, por sí sola, no es suficiente: debe ser regulada, controlada. El método cobra de este modo relevancia y carácter de guía (él conduce de los particulares a las causas y principios universales): se hace imprescindible hallar el modo correcto de llevar las experiencias a cabo, interpretarlas y difundirlas. Sin embargo, el concepto de experiencia no tenía un significado unívoco en esta época. En este sentido, Hobbes (véase a este respecto su polémica con Boyle acerca de la *filosofía experimental*) o Descartes son ejemplos de *filósofos naturales* que mantuvieron una concepción ciertamente tradicional de experiencia, pues la entendieron como observación de aquello que sucede de forma natural –sin control experimental, diríamos hoy–. Francis Bacon, por su parte, patrocinó una interpretación diferente: según su planteamiento, era necesario el registro de una historia natural debidamente compilada que incluyera tanto lo que en la naturaleza se da ordinariamente como lo que en ella es ocasional o "monstruoso", además de los experimentos producidos artificialmente. Puede pues apreciarse con meridiana claridad la referida pluralidad emergente en aquella reconstrucción del edificio del conocimiento emprendida por los *filósofos naturales* del siglo XVII.

No obstante, no es en esta pluralidad en lo que aquí nos interesa incidir, sino precisamente en lo contrario, esto es, en la presencia de significativas notas comunes en los autores de la época de aquella seminal intersección entre filosofía y ciencia. Por lo que al origen de programas filosóficos naturalistas respecta, cabría entre dichas notas destacar las siguientes: la ciencia se presenta como una filosofía rigurosa, el titubeo especulativo es mirado con recelo, el catálogo ontológico canónico empieza a purgar elementos que no resisten el paso a través del cedazo de la experiencia y los temas filosóficos tradicionales empiezan a analizarse dentro de este renovado contexto intelectual. Lo hu-

mano, como hemos puesto de relieve mediante el ejemplo de Hobbes, cae dentro del ámbito mecanicista de la nueva ciencia y la mente comienza a ser tratada como una realidad conceptualizable en dichos términos.

> Es natural que con el nacimiento de la ciencia moderna no se resistiera la tentación de cosificar la mente. En el fondo, se trataba de objetivarla, a fin de hacer de ella objeto de la ciencia. (Hierro-Pescador, 1997: 36)

Esta concepción mecanicista de la mente de los *filósofos naturales* anteriores a Newton se halla inserta en una concepción de lo mecánico completamente ajena a la noción de campo: todo efecto físico ha de ser resultado del impacto de una partícula material sobre otra. Este marco teórico se vería ampliado con el advenimiento de la física newtoniana, que complementara al mecanicismo cartesiano al integrar en su seno, junto a la efectividad causal los impactos entre entidades materiales, la de unas fuerzas incorpóreas que vinieran a ofrecer una nueva carta de ciudadanía al dualismo interaccionista al abrir la puerta a la especulación acerca de la existencia de fuerzas mentales en algún sentido análogas a las magnéticas o las gravitacionales. La mente inmaterial volvía a poder ser concebida como instancia capaz de producir efectos en el mundo físico y así, la física del siglo XVIII no puso trabas a la especulación filosófica acerca de fuerzas mentales no físicas pero capaces de producir efectos físicos, unas fuerzas mentales análogas a las fuerzas postuladas por la física newtoniana, incorpóreas pero eficaces causalmente.

No obstante, también en el siglo de los primeros newtonianos puede rastrearse la continuidad de planteamientos naturalistas como los inaugurados por Hobbes, por ejemplo, en la *ciencia del hombre* que Hume reivindicara en la introducción al *Treatise*. En el marco de dicha ciencia, presentada como una suerte de centro articulador del sistema completo de las ciencias, la esencia de la mente, sus propiedades, habrían de ser investigadas en base a la experimentación y la observación, según el modelo de las ciencias empíricas. De este modo, Hume sienta las bases de una nueva concepción mecanicista de lo mental, articulada en torno a sus principios asociativos y sus nociones de impresión e idea. En Hume, en continuidad con la sugerida evanescencia de las fronteras entre ciencia y filosofía dentro de un incipiente programa naturalista, "no se da una separación nítida entre ciencia y filosofía" (Guerrero del Amo, 2000: 74) y las cuestiones empíricas no quedan excluidas del ámbito de la filosofía. No encontramos una brecha que separe radicalmente la investigación empírica de la filosófica, sino que, al con-

trario, introduce Hume en la esfera de la filosofía el método experimental. El rechazo humeano de la metafísica se prolongará asimismo en la tradición naturalista. Se trata de un aspecto de la producción humeana que conecta su defensa del método experimental con su propósito de establecer unos límites fiables del conocimiento, cuestiones que hallamos perfiladas en la primera sección de *An Enquiry Concerning Human Understanding*.

Ya en el siglo XIX un importante hito en la gestación del naturalismo filosófico contemporáneo y, asimismo, en la reflexión científica y filosófica acerca de la mente llegaría de la mano del unánime consenso acerca de las leyes de conservación de la física cuando, a mediados de siglo, la conservación de la energía se convirtiera en un principio básico de la física. La carta de ciudadanía que a la mente inmaterial concedieran las referidas fuerzas de la física newtoniana le sería así nuevamente arrebatada. A ningún *filósofo natural* le estaba ya permitido postular entidades ajenas a dicho principio al especular acerca de peculiares fuerzas mentales no sujetas al gobierno de leyes deterministas articuladoras de un cosmos que no tolera variación en su cantidad total de energía. Esta restricción a la especulación filosófica derivaría en el a día de hoy vigente principio del cierre causal del mundo físico, según el cual todo efecto físico, esto es, todo evento físico causado, es necesariamente resultado de causas físicas y tiene de hecho una causa física completa. Que la causa sea completa significa que quedan contemplados los habitualmente denominados factores causales, con lo cual, cabría reformular el principio señalando que el mismo sostiene que "ninguno de los factores causales involucrados en la producción de un efecto físico es no físico" (Vicente, 2001: 4). Ningún insumo causalmente eficaz y exógeno al espacio-tiempo de nuestro universo es posible según este principio.

Finalmente, consideramos que el último acontecimiento intelectual decisivo para la conformación del naturalismo contemporáneo tuvo lugar en la segunda mitad del siglo XIX con el desarrollo de la biología evolucionista. Es bien sabido que, a pesar de precedentes como la *Philosophie Zoologique ou Exposition des Considérations Relatives à L'histoire Naturelle des Animaux* que Jean-Baptiste-Pierre-Antoine de Monet, chevalier de Lamarck publicara en 1809, o los especulativos *Vestiges of the Natural History of Creation* que Robert Chambers publicara anónimamente en 1844, el centro de gravedad de este desarrollo se encuentra en la publicación de *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* el 24 de noviembre de 1859. Este texto contiene "la idea incuestiona-

blemente más importante que jamás haya concebido la mente humana" (Romanes, 1892: 257τ): la de la selección natural. Dicha idea, simultánea e independientemente acariciada por Alfred Russel Wallace, sigue siendo aceptada como el núcleo de la explicación de la diversidad biológica y no ya como el único motor de la evolución de las especies, pero sí como el decisivo. Según la misma, en resumidísimas cuentas y en una versión contemporánea, el éxito reproductivo diferencial de las distintas variantes genéticas[3] presentes en una población está a la base del proceso evolutivo: los fenotipos de los organismos capaces de dejar más descendencia preponderarán en sucesivas generaciones dando, eventualmente, lugar a procesos de especiación. La aplicación de los planteamientos darwinianos al estudio de lo mental no se hicieron esperar. El propio Darwin emprendió dicha aplicación transcurrida apenas una década desde la publicación de su obra capital, en los capítulos tres a cinco de la primera parte –de la segunda edición: el matiz es importante, dada la diferente ordenación de los materiales del segundo y cuarto capítulo entre la primera y la segunda edición– de *The Descent of Man, and Selection in Relation to Sex*, de 1871/1882, y en *The Expression of the Emotions in Man and Animals*, de 1872. Ya en la primera edición de *The Origin of Species* Darwin apuntaba: "Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation" (Darwin, 1859: 397), una idea que, espoleado por Chambers, había avanzado ya Herbert Spencer en sus *Principles of Psychology*, de 1855. La aplicación de los planteamientos darwinianos al estudio de lo mental sería inicialmente retomada por el feraz polímata y primo de Darwin Sir Francis Galton, así como por los pragmatistas americanos, pero de sus frutos siguen viviendo en uno u otro sentido todas y cada una de las disciplinas dedicadas a la investigación del sistema nervioso, la mente y la conducta.

### 1.2._El naturalismo contemporáneo

Hasta aquí nos hemos ocupado de los antecedentes de la perspectiva naturalista: planteamientos que sin duda cabe vincular directamente con eso que hoy denominamos naturalismo, pero a los que no cabe denominar naturalistas en sentido contemporáneo. La acepción contemporánea del término proviene de la filosofía norteamericana de la

[3] A pesar de que Mendel conocía la obra de Darwin, que estudió en una edición alemana de 1863, y a pesar de que una de las cuarenta copias de su "Versuche über Pflanzen-Hybriden" que encargara imprimir para su envío a las principales figuras de la biología europea fuera, muy probablemente, enviada a Darwin, éste, por así decir, no acusó recibo y dejó en blanco el capítulo sobre los mecanismos de la herencia.

primera mitad del siglo XX. Autores como John Dewey, Ernest Nagel, Sidney Hook, John Herman Randall o Roy Wood Sellars se declararían en aquel entonces naturalistas y dotarían al término de contenido en un intento por aproximar la reflexión filosófica a la investigación científica mediado por la intuición de que todo lo existente debe ser clasificado bajo la signatura "natural", incluyendo la mente humana, que, por tanto, habrá de ser desentrañada del mismo modo que el resto del *mundo natural* (locución, según lo antedicho, pleonástica): echando mano del método científico. La mente y su estatuto ontológico sería, pues, un aspecto central de este incipiente naturalismo contemporáneo, como evidencia la publicación de *Naturalism and the Human Spirit*, en 1944. Esta obra colectiva, editada por Yervant Hovhannes Krikorian, contenía textos de las más destacadas figuras del naturalismo filosófico americano[4] y fue concebida –según Kim (2003: 86) o Gasser (2007: 4)– como el manifiesto de dicho movimiento. La centralidad que para los filósofos naturalistas de la primera mitad del siglo XX tuviera la cuestión de la mente humana y el modo de abrir el camino hacia una comprensión científica de la misma se hace enteramente evidente en este volumen. En la reseña que Arthur E. Murphy hiciera del mismo podemos leer: "Starting from the acknowledged achievements of scientific inquiry so far, the 'naturalists' intend to show that these same methods, or others essentially 'continuous' with them, are adequate also to those aspects and dimensions of 'the human spirit' which in the past have often been held on philosophical grounds to transcend the methods and aims of science" (Murphy, 1945: 405). Dentro del marco de este naturalismo filosófico de la primera mitad del siglo XX, el principio del cierre causal del mundo físico seguía tan vigente como en el siglo anterior, haciéndose patente su influencia en la reflexión acerca de lo mental en el incipiente fisicalismo de mediados de siglo (como ejemplo de este influjo del principio del cierre causal en la nueva ontología naturalista de medidos del siglo XX, vid. Feigl, 1958; Oppenheim & Putnam, 1958). Sirve para ilustrar el carácter medular de este principio en los orígenes del naturalismo contemporáneo el modo en que Sellars padre (Sellars, 1927)

[4] Merece la pena mencionar dos ausencias: la de George Santayana y la de Roy Wood Sellars. La ausencia de Sellars responde al hecho de que este texto de 1944 no es *el* manifiesto de aquel incipiente naturalismo filosófico contemporáneo, como sugieren Kim o Gasser, porque de hecho hubo dos manifiestos: el de cuño pragmatista al que aluden Kim y Gasser, que incidiera en una interpretación metodológica del naturalismo, y otro centrado en cuestiones de tipo ontológico y encarnado en *Philosophy for the Future. The Quest of Modern Materialism*, que Sellars editara junto con Marvin Farber y Vivian Jerauld McGill en 1949. Tal y como Nunziante (2013: 42 y ss.) ha señalado, sería precisamente el debate que entre Sellars y Hook tuviera lugar acerca del estatus de la experiencia subjetiva el eje que articulara la ruptura entre ambos grupos de filósofos naturalistas. No obstante, dados los propósitos de la presente exposición, no entraremos en detalles ni echaremos mano de esta distinción entre dos formas de naturalismo en la filosofía norteamericana de la primera mitad del pasado siglo XX.

presenta la noción de naturaleza como idéntica a la de realidad y, al tiempo, como el sistema espacio-temporal-causal completo y autosuficiente que estudian las ciencias naturales. Nada pareció amenazar en las décadas subsiguientes la vigencia de esta idea dentro de los círculos filosóficos naturalistas –de hecho, cincuenta años después de Sellars, David Armstrong la formula en términos prácticamente idénticos: "Nature, the spatio-temporal system, is a causally self-enclosed system" (Armstrong: 1978a: 153).

Siendo nuestro objetivo en este apartado el de perfilar los contornos generales del naturalismo en filosofía, no podíamos dejar de ocuparnos brevemente del programa naturalista que disfrutara de mayor resonancia durante la segunda mitad del pasado siglo XX, el de Willard van Orman Quine. El naturalismo quineano se presenta primariamente como un proyecto de naturalización de la epistemología. A su vez, epistemología y filosofía aparecen en el marco del mismo como nociones apenas diferenciables, casi coextensivas. No intentaremos abordar aquí un minucioso análisis de la obra de Quine, sino sólo ofrecer algunas pistas generales de las cuales servirnos para ilustrar someramente la forma de su programa naturalista, un programa en cuyo ascendiente y pujanza no consideramos necesario incidir y al que, entendemos, cabe apuntar como hito central e ineludible punto de referencia de los planteamientos filosóficos contemporáneos en los que la reflexión filosófica y la investigación científica no se presentan como pertenecientes a ámbitos claramente diferenciados, sino como empresas apenas distinguibles que avanzan conjuntamente abriendo un mismo camino en su intento por comprender el mundo.

En *Pursuit of Truth* (Quine, 1990), síntesis de su pensamiento publicada tras seis décadas de actividad académica, Quine se pregunta acerca de la posibilidad de nuestro conocimiento del mundo, se pregunta cómo es dicho conocimiento posible, y avanza que es la ciencia la que debe ofrecer respuesta a tal pregunta. De este modo, la cuestión central de la epistemología se convierte en objeto de una disciplina científica: la psicología. Esta idea arraiga en la tesis del continuo entre filosofía y ciencias, según la cual no existe un hiato entre ambas. Dicha tesis conforma, junto con el abandono de la pretensión tradicional de construir una *filosofía primera*,[5] anterior a las ciencias particulares y rectora de las mismas, el núcleo del planteamiento quineano. Es claro, pues, que aquí las fronteras entre ciencia y filosofía se muestran nuevamente evanescentes, diluyéndose en un proceso en el cual la primera alcanza y resuelve problemas tratados tradi-

[5] Recientemente Penelope Maddy (Maddy, 2007) ha jugado con esta noción presentando al filósofo naturalista como un prudente "second philosopher".

cionalmente por la segunda, un proceso que no debiera arredrar al filósofo, al que podemos entender que en este contexto no se le arrebatan ni su capacidad para el esclarecimiento conceptual ni –en vista de los análisis del propio Quine– la posibilidad de ofrecer perspectivas metacientíficas. Así, filosofía y ciencia, en un extraño movimiento, se incluirían mutuamente, dado que la epistemología aparecería como una rama de la psicología mientras la filosofía se encargaría de criticar, elucidar y esclarecer concepciones generales y conceptos fundamentales. Respecto de la tesis del continuo es interesante hacer notar –volviendo sobre la anfibológica nomenclatura de la que hiciéramos mención al tratar de los *filósofos naturales* del siglo XVII– que Quine (1979: 191) subraya que algunos de los intelectuales que hoy reverenciamos como grandes filósofos fueron de hecho científicos, pensadores inmersos en la búsqueda innominada –podríamos decir– de una concepción coherente y organizada del mundo. La tesis del continuo y la del abandono de la metafísica o *filosofía primera*, los dos primeros puntos de la agenda naturalista quineana, se encuentran, de hecho, relacionados entre sí: no existe un saber privilegiado anterior o exterior a las ciencias particulares, de modo que ciencia y filosofía no mantienen una relación basada en la autoridad de una sobre otra, sino que se hallan implicadas en una tarea convergente dentro de la cual la única autoridad es la realidad misma, la experiencia, ante la cual se encuentran juntas y en idéntica situación las ciencias y la filosofía. Ni unas ni otra reclaman una posición privilegiada ni se entienden a sí mismas como fundadoras o fundamentales. La epistemología no puede entenderse, desde la óptica naturalista de Quine, como una labor desvinculada de la ciencia. Dicha rama de la filosofía fue concebida tradicionalmente como la disciplina encargada de la justificación y fundamentación del conocimiento científico, con lo cual, un punto habitual de discusión del naturalismo quineano ha venido siendo el de si resulta o no legítimo que la justificación del conocimiento científico haga uso de resultados científicos. Sin embargo, cabe aducir contra esta objeción de circularidad que el filósofo naturalista no parte de la necesidad de un saber fundante y, por añadidura, niega que sea posible o necesaria una fundamentación última del conocimiento, es decir, una epistemología trascendental, externa a la ciencia. En línea con esta objeción de circularidad, algunos han encontrado en la propuesta quineana una llamada al abandono de la normatividad en epistemología (vid., v. g., Putnam, 1982; Kim, 1988b). Constituye poco menos que un lugar común señalar que, dado que programas naturalistas como el de Quine o el de las teorías evolucionistas del conocimiento se ven abocados a centrarse en la consideración de cuestiones de hecho –en particular, en la descripción y explicación de

las creencias y su génesis–, queda fuera de su alcance toda posibilidad de aproximación a extremos nucleares de las tradicionales teorías del conocimiento, como la normatividad, la certeza, la validez o la racionalidad. Señalemos, no obstante, que Quine sugirió de hecho vías para ofrecer pautas epistemológicas normativas –desembocando en su tecnología de la predicción de estímulos sensoriales (Quine, 1990)– y, lo que es más notable, que llegó a ellas movido por su afán de evitar el escepticismo y el relativismo. Pero dichas pautas no excedían la recomendación de determinados derroteros aconsejables de cara a alcanzar objetivos no discutidos, pues en el planteamiento quineano los fines de la ciencia están ya inscritos en el propio terreno de juego científico y no se presentan así como objetos de discusión racional. En cualquier caso, la objeción antinaturalista en este punto habría de ser planteada de tal modo que despejara toda hesitación posible, tanto acerca del tipo de normatividad a la que la misma alude como acerca del modo en que la tradicional epistemología trascendental se hace necesaria para alcanzarla mientras que una atención crítica y reflexiva a las condiciones empíricas de la génesis del conocimiento se mostraría incapaz de dilucidar valores epistémicos y elucidar modos y prioridades en el acceso a los mismos. En cualquier caso, no es éste, desde luego, el lugar para disertar acerca del modo correcto de entender y elaborar una epistemología naturalizada, como tampoco lo es para la discusión del modo correcto de interpretar a Quine. Apuntemos, con todo, que no faltan quienes, como Susan Haack, han aludido a la falta de concreción de la propuesta de Quine como el motivo de su éxito. Según la filósofa británica, Quine llega a defender formas radicales, intermedias y moderadas de epistemología naturalista dentro de un mismo párrafo (Haack, 2009: 167 y ss.). Así, a nadie debiera extrañar que a prácticamente cualquiera le sea dable encontrar en Quine a su naturalista de cabecera, pues de hecho estaría defendiendo, desde la perspectiva de Haack, prácticamente cualquier suerte de naturalismo, desde versiones fuertes que desterrarían el proyecto tradicional de la epistemología al completo y sustituirían sus desencaminadas indagaciones por investigaciones bien acotadas dentro del ámbito de las ciencias de la cognición, a versiones suaves y más bien inocuas según las cuales la epistemología, sencillamente, no es una disciplina completamente apriorística, sino que mantiene una cierta relación de continuidad con las ciencias naturales, cuyos resultados pueden resultar relevantes para el quehacer del epistemólogo. Como indicábamos, no es nuestro objetivo en este punto el de contribuir a la exegética quineana. Nos basta con presentar su programa como genitor de proyectos filosóficos contemporáneos en los que las demarcaciones entre departamentos de filosofía y ciencias aparecen desdibujadas.

La prescripción del abandono de la metafísica aparece en la propuesta de Quine como rechazo de una perspectiva consolidada sobre el papel y el lugar de la epistemología. Entre los siglos XVII y XVIII, en una tradición que podemos remontar a Descartes y ver culminar en Kant, la epistemología, la moderna teoría del conocimiento, había desplazado a la metafísica como disciplina fundamental –tierra firme sobre la que se yergue el edificio todo del conocimiento humano–, pero, en este proceso, la epistemología desbanca a la metafísica para convertirse ella misma en metafísica, para convertirse ella misma en *filosofía primera*, la disciplina fundamental, anterior y exterior al resto del edificio del conocimiento humano, al cual ofrece fundamento y justificación. Quine, tomando buena nota de la incapacidad de contemporáneas propuestas epistemológicas empiristas (tiene en este punto en mente el proyecto del Círculo de Viena, y en concreto la carnapiana reconstrucción fenomenista del significado de los conceptos científicos) para alcanzar sus ambiciosos objetivos fundamentalistas, aconseja desechar esa concepción según la cual la epistemología es una disciplina que elabora a priori y con independencia de teorías y resultados científicos fundamentación y justificación del conocimiento científico.[6] De este modo, el barco de la famosa imagen que Quine tomara de Neurath se reconstruye desde dentro, en el mar y estando a flote, y la justificación que las teorías científicas requieren es la justificación que la misma ciencia exige a las teorías científicas: la epistemología como disciplina apriorística garante de fundamentación no está presente en el cuadro quineano. Abandonada esta pretensión, Quine invita a los epistemólogos a olvidar semejantes prerrogativas fundamentalistas, autónomas y aprioristas y contribuir en la empresa de comprender cómo se gesta realmente el conocimiento científico, una invitación que no se prolonga en rechazo de tareas epistemológicas de corte tan tradicional como el análisis del método científico o la clarificación de conceptos básicos o de alto nivel, pero, eso sí, ahora desde dentro de las propias fronteras de la ciencia: los límites de la epistemología son los límites de la ciencia, es decir, no existe base más sólida para la fundamentación de la certeza científica que el propio método científico.

El naturalismo de Quine puede, en definitiva, interpretarse de dos modos. Según la primera lectura Quine prescribe trascender la comprensión de la epistemología como una *filosofía primera* capaz de fundamentar el conocimiento científico desde fuera del

[6] "Our Second Philosopher freely acknowledges one poignant aspect of the human condition: we can't step outside our system of beliefs and methods and justify them from an external perspective; the only perspective we can occupy is our own" (Maddy, 2007: 35).

mismo. De acuerdo con la segunda, defiende que la epistemología debe practicarse dentro del marco de la ciencia natural, pues su objeto es un producto natural: el conocimiento humano. Se trata de diferentes interpretaciones en tanto cada una de ellas enfatiza distintos aspectos del programa de Quine, pero no se trata de lecturas mutuamente excluyentes pues, como puede observarse, cada una de ellas se centra en una de las dos tesis que hemos presentado como los dos elementos interrelacionados sobre los que pivota el naturalismo quineano: la tesis del continuo y el abandono de la *filosofía primera* como garante de todo conocimiento, como fundamentación extrínseca del resto del conocimiento.

Como indicábamos, no pretendemos con esta somera presentación del naturalismo quineano resultar exhaustivos (es obvio que para ello habríamos de entrar en numerosos aspectos de su filosofía del lenguaje, la lógica y la ciencia a los que no hemos aludido), sino sólo preparar el terreno para la introducción del marco naturalista al que nos adheriremos. A este fin, es suficiente la escueta elocuencia con que, en un puñado de palabras, presenta Daniel Dennett el que también nosotros consideramos el núcleo de la perspectiva naturalista en filosofía:

> Quine was a naturalist, and for him philosophy was continuous with science. It was just the sort of most abstract and most meta part of the scientific enterprise. It was science criticism, science enabling, trying to put the world together with science, and I just completely bought into that. (Dennett, 2007: 3).

Para Dennett, y para el resto de los filósofos alineados en este punto con Quine, la filosofía no se ubica en un lugar privilegiado respecto de las ciencias; siquiera se encuentra, en cuanto forma de conocimiento y crítica del conocimiento, claramente separada de ellas. La filosofía naturalista no se presenta como alguna suerte de púlpito externo a nuestros sistemas de creencias, métodos y teorías en el cual se hallen a nuestra disposición formas a priori de conocimiento y desde el cual nos quepa organizar normativamente el edificio completo del saber humano. Científicos y filósofos juegan al mismo juego y en el mismo equipo en su prosecución de una forma rigurosa y coherente de acceso epistémico a la realidad. Ningún tribunal más allá de la propia evidencia racionalmente estipulada puede alzar la voz contra la ciencia, pero, además de la labor crítica que la ciencia misma desarrolla, filósofos con la adecuada preparación pueden aportar perspectivas fructíferas mediante reflexiones críticas informadas, cuidadosos estudios de los fundamentos de las disciplinas o especialidades que sean el caso y escrupulosos exámenes de los desarrollos producidos en el seno de las mismas. La frontera entre

ciencias y filosofía, en definitiva, es algo que no acierta a ver el naturalista: "philosophy at its best and properly conceived is continuous with the empirical sciences" (Churchland, 1986: 2).

## 2._Nuestro marco naturalista

En este apartado definiremos y defenderemos la postura filosófica en que se basa nuestra interpretación de las derivas por las que en las últimas décadas ha venido transcurriendo el debate en torno al estudio de la conciencia. «Naturalismo» es la etiqueta usualmente empleada para calificar y clasificar posturas similares a la que defenderemos, pero esta etiqueta ha sido indiscriminadamente utilizada en filosofía de la mente, de tal modo que en una u otra ocasión se ha denominado "naturalista" a la práctica totalidad de la producción contemporánea en dicha área. Así las cosas, habremos de precisar en lo sucesivo qué entendemos aquí por naturalismo para poder después aclarar en qué sentido entendemos que la mente y la conciencia no obligan a abandonar el marco teórico naturalista, sino que pueden y deben ser cabalmente integradas en el mismo. Pero, ¿cómo perfilar dicho marco? Será la respuesta a esta pregunta lo que nos ocupará en lo que resta de este apartado.

Con mucha frecuencia y excesiva laxitud se define el naturalismo filosófico por contraposición al supernaturalismo o sobrenaturalismo filosófico.[7] Así, se afirma que un filósofo naturalista no puede apostar por la existencia de entidades que excedan el orden natural, queriendo con «natural» significar todo aquello que se halla sometido a causas y leyes naturales siendo así en principio pasible de descripción, explicación y predicción haciendo uso de las herramientas y métodos empleados por las ciencias naturales. Así definido, el naturalismo no pasa de constituir una opción ontológica cientista en cuyas coordenadas no hay cabida para dioses, ángeles o almas, una opción ontológica adornada por un corolario metodológico de gran relevancia por lo que al estudio de la mente toca: nuestra investigación de la naturaleza humana ha de prescindir de toda apelación a órdenes trascendentes, prosapias divinas, esencias misteriosas, almas y demás entelequias y atenerse a datos y procedimientos análogos a los utilizados en el estudio de la quimiosíntesis, la autólisis, la partenogénesis o el vulcanismo. La laxitud de esta forma

[7] De hecho, Stroud (1996: 54) llega a sostener que sólo en el anti-supernaturalismo encuentra el manoseado término «naturalismo» una justificación y una acepción legítima.

de definir el naturalismo filosófico se hace patente, por ejemplo, en el hecho de que la mayoría de los filósofo contemporáneos, independientemente de su mayor o menor cercanía con prácticamente cualquier tendencia o tradición filosófica, no tendría inconveniente en ser tenido por naturalista si con ello no pretendiera indicarse nada más allá de determinada forma de negación de la existencia de entidades sobrenaturales como las referidas. La opción ontológica a la que aludíamos levanta pues pocas ampollas hoy día. Bien distinto es el panorama por lo que toca al apuntado corolario metodológico: la idea de que hayamos de acudir a las ciencias de la naturaleza para estudiar –parafraseando a Teilhard de Chardin– el "fenómeno humano" despierta menos simpatías y topa con todo tipo de oposiciones y resistencias –no así la de que en situación pareja nos encontramos respecto del "fenómeno canino", anotemos–. No entraremos en detalles por lo que a ellas se refiere, dado que sólo nos interesa aquí destacar la vaguedad de las definiciones de «naturalismo» que disfrutan de mayor difusión en la actualidad y tratar de matizarlas para dotar así de contenido al término. En este sentido, la noción de naturalismo es actualmente utilizada en filosofía con una extensión quizá excesivamente amplia y, tal y como Papineau (2007) ha puesto de relieve, un significado no particularmente preciso. Así las cosas, la reciente polémica que acogieran las páginas de opinión del New York Times entre Alex Rosenberg y Timothy Williamson es totalmente comprensible: el marco naturalista se presenta de forma tan vaga que a nadie se le hace extraño que éste (Williamson, 2011), ateniéndose a una definición no mucho más precisa que la expuesta, pretendiera mostrar cómo puede el naturalismo frisar el dogma, ni que aquél (Rosenberg, 2011) respondiera contundentemente presentando razones para mantener una actitud de reserva ante quienes hablan de la hermenéutica, la crítica literaria, el posestructuralismo o la deconstrucción como formas de conocimiento.

Entendemos, por nuestra parte, que el núcleo de una propuesta filosófica naturalista consiste, efectivamente, en la negación de la tesis supernaturalista, pero entendemos asimismo que ésta ha de definirse con cierta cautela. Según dicha tesis, existirían agentes o fuerzas ajenas o externas al mundo natural, agentes o fuerzas que, además, podrían tener algún efecto causal en el curso de los acontecimientos que en éste tienen lugar. Se trata de una tesis ontológica adornada por un corolario epistemológico que, desde nuestro punto de vista, es lo que resulta verdaderamente comprometido. Según el mismo, la creencia en dichos agentes y sus poderes causales no puede hallar sustento en ninguno de los métodos de los que nos es dable servirnos para obtener conocimiento de forma

fidedigna en cualquiera de las ramas del saber.[8] Del mismo modo que difícilmente cabe hallar justificación para la más bien poco definida tesis de acuerdo con la cual todo es natural, difícilmente cabe hallarla para la tesis contraria. Un prudente voto de humildad epistémica habría de prevenirnos de pronunciarnos en uno u otro sentido. La fuerza de la negación naturalista de la tesis supernaturalista se encuentra, pues, en el modo en que el supernaturalista se siente exento de hacer dicho voto, como evidencia el señalado corolario epistemológico. Desde luego, nadie parece hallarse en disposición de sentenciar acerca del carácter o la naturaleza de todo lo existente, pero la navaja de Ockham sigue inclinando al filósofo epistémicamente humilde a no comprometerse con tesis dogmáticas según las cuales X existe porque X existe y no cabe buscar justificación para nuestra creencia en su existencia más allá de nuestra intuición o nuestros sentimientos. Hasta aquí nuestra escueta defensa del núcleo de cualquier programa naturalista en filosofía. A pesar de tratarse de una defensa excesivamente sucinta, lo en ella contenido no deja de resultar decisivo por cuanto presenta al naturalismo como un planteamiento filosófico con obvias implicaciones ontológicas, pero de cuño epistemológico. ¿Por qué? Porque en ella se hacen palmarios los motivos del naturalista: no está dispuesto a creer dogmáticamente en entidades sobrenaturales, pero no porque sufra alguna clase de alergia, sino porque las mismas se muestran huidizas, inaccesibles haciendo uso de cualesquiera métodos de indagación racional de la realidad. Richard Lewontin ha comentado en este sentido que "el eminente académico kantiano Lewis Beck solía decir que quien puede creer en Dios puede creer en cualquier cosa" (Lewontin, 1997: 31τ). El naturalista no necesita, desde la óptica de nuestra escueta defensa del núcleo del naturalismo filosófico, ostentar tal o cual militancia ontológica, no necesita patrocinar esta o aquella tesis según la cual lo único que existe es esto o aquello y tiene tal carácter o tal otro: basta con que afirme que nuestras prácticas epistémicas, sumadas a la debida parsimonia gnoseológica y a una humilde y cabal sindéresis, nos permiten hablar con cierta seguridad ontológica de huracanes, pero no así de Eolo o los Anemoi. No obstante, como avanzábamos, esta escueta defensa atiende sólo a una caracterización restringida de «naturalismo» que no pasa de constituir una definición mínima de dicho término y que, así, exige ulteriores esfuerzos por ampliar y matizar la noción en cuestión.

[8] Es obvio que el recurso, tácito o explícito, a este corolario epistemológico puede servir para salvaguardar la ontología dualista. Así, por ejemplo, John Eccles, en su debate con Derek Denton, acude al mismo cuando arguye que, según su dualismo interaccionista, lo mental y lo neuronal interactúan de la manera más sutil, pero que "con *ninguno de nuestros trucos* podemos separarlos" (Denton, 1993: 103τ). [Las cursivas son nuestras].

Más allá de la negación del supernaturalismo y la defensa de una ontología "cientista" –una ontología naturalista que podría presentarse de muchas maneras, pero que, según lo expuesto, habría de ser comprendida como parasitaria del núcleo epistemológico del naturalismo: nuestra concepción de la realidad no puede incluir entidades o fuerzas al alcance de, exclusivamente, formas a priori de teorización, intuiciones trascendentales injustificables o filosofías primeras de cualquier clase–, lo definitorio de la postura naturalista en filosofía consiste en una concepción de la filosofía misma, una concepción según la cual, más allá del hecho obvio de que "nunca ha sido posible hacer filosofía al margen de la ciencia" (González, 1989: 239), existe tal continuidad entre la indagación científica y la filosófica que delimitarlas o deslindarlas resulta extremadamente complicado, pues no hay una brecha entre ambas, no existe un método filosófico heterogéneo respecto del científico, no existe una forma filosófica de investigar la realidad radicalmente diversa de ésa que las diferentes disciplinas científicas emplean, no existen formas a priori de investigación, no existe una *filosofía primera*, no existen ni procedimientos ni productos cognitivos propia o esencialmente filosóficos, no existe una peculiar forma de conocimiento que quepa adjetivar como "filosófica", "no hay ninguna puerta de acceso independiente a la verdad para los filósofos" (Reichenbach, 1949: 310τ). "Ciencia y filosofía forman un continuo. La filosofía es la parte más global y reflexiva del continuo, el escenario de las discusiones que preceden y siguen a los avances científicos" (Mosterín, 2003a: 212; 2013a: 22). La acumulación de negaciones en los últimos renglones no pretende dar pábulo al vaticinio preceptivo según el cual la filosofía deberá desaparecer diluyéndose en esta o aquella rama de las ciencias, sino, al contrario, a la idea de que "basta con renunciar a métodos que reclamen una supuesta autoridad epistémica basada en intuiciones incontrastables" (Diéguez Lucena, 2014: 26). "La ciencia se ha convertido en una inexcusable compañera de cualquier otra actividad intelectual" (Moya, 2010: 21), y particularmente de cualquier forma de actividad intelectual que, como vino entendiéndose que hace la filosofía, se pone por meta la purga y extensión de nuestra comprensión por medio del análisis crítico, siendo así que los datos empíricos, los resultados experimentales y las propias teorías científicas no son un mero adorno a la especulación filosófica que quepa integrar de un modo u otro en el curso de la misma: son materiales ineludibles para la reflexión filosófica. Llamamos filosófica a esta reflexión no porque algo le sea peculiar: el científico no ejercita una supuesta Racionalidad 1 en el análisis de los señalados resultados y teorías diferente de una supuesta Racionalidad 2 que al filósofo le cupiera desplegar en dicho contexto. Se trata, senci-

llamente, de profesionales con formaciones diversas que enfocan, en último término, unos y los mismos problemas, acaso, desde diferentes ángulos dadas las condiciones que han conducido a sus respectivas disciplinas al planteamiento de los mismos.

Nuestra adquisición de conocimiento es un proceso natural que tiene lugar en el contexto de una historia: la de la relación entre nuestra especie y su entorno. No podemos salir fuera de esta historia para explicar aquel proceso desde un fingido *Punctum Archimedis*: no hay cátedra o título universitario que ofrezca la posibilidad de hacerlo. El plural curso de la conformación de métodos destinados a propiciar el apuntado proceso constituye la médula de la referida historia. Encauzar ese curso es una tarea que difícilmente cabe concebir que pueda ser llevada a término desde fuera del mismo: a nadie le es dable elevarse por encima de aquella historia alzando el vuelo sobre nuestros métodos y sistemas de creencias para juzgarlos, ordenarlos y justificarlos desde la hipotética perspectiva externa que semejante taumaturgia escapista brindaría. Muchos se ven tentados al ser expuestos a esta clase de razonamiento a firmar partes de defunción como el que recientemente redactaran Stephen Hawking y Leonard Mlodinow: "la filosofía ha muerto" (Hawking & Mlodinow, 2010: 5τ). Parecen entender que si a la filosofía no se le puede reservar una especialísima –aunque asimismo indeterminadísima– parcela, autónoma respecto de las ocupadas por las diversas disciplinas particulares, no se le puede reservar ninguna en absoluto. "Si la clase de conocimiento que el naturalista concibe como fidedigna –se dicen– es ésa que rezuma del cedazo del método científico, lo que el naturalista está pidiendo a los filósofos es que recojan sus bártulos y dejen que las ciencias se encarguen de todo". Muy al contrario, cabe, por una parte, preguntar si no es "acaso filosófica la herramienta básica que utiliza la ciencia, que es el método científico, [y si no lo son] (…) la abstracción y la formulación rigurosa de las hipótesis que hacen los científicos" (Mora Teruel, 2007: 47), y, por otra, aseverar que la de comprender, que es precisamente la tarea que las ciencias y la filosofía comparten,[9] es una tarea que ha venido beneficiándose de la pluralidad de perspectivas, y muy probablemente, por ejemplo, las apreciaciones acerca del vínculo entre determinados resultados y determinada hipótesis que un especialista en la materia que sea el caso pudiera realizar diferirán e, idealmente, se beneficiarán de su contrate con las que un teórico instruido en la misma que, por añadidura, haya recibido formación en las disciplinas del análisis

[9] Una tarea cuyo ser compartida se halla si cabe más justificado, como advierte Alva Noë, cuando de lo que se trata de es comprender la conciencia o, sencillamente, nuestra propia naturaleza (Noë, 2009: xv del original, 18 de la traducción, que tergiversa en este punto el sentido del original al verter "our own nature" como "la propia naturaleza").

formal del razonamiento y la argumentación pudiera plantear. De este modo, al filósofo naturalista interesado en fomentar la precisión, amplitud y penetración de nuestra comprensión en cualquiera que sea el área le cabe participar activamente en ella de muy diversos modos, ninguno de los cuales le será exclusivo y el crucial entre los mismos residirá en hacer con las hipótesis, teorías y métodos lo que ya hacen los científicos: tratar de refinarlos buscando sus puntos débiles y haciendo espacio para que otros mejores vengan a sustituirlos. Cuando los datos y los planteamientos teóricos están publicados llega la parte verdaderamente científica, la propiamente filosófica: la de reflexionar y criticar tratando de contribuir a que las piezas encajen del mejor modo. No hay ningún motivo de principio para que el filósofo deje de jugar a este juego, el decisivo, pero hay muchas otras tareas adecuadas a su formación y necesarias para el avance de las disciplinas científicas, como el esclarecimiento conceptual mediante el análisis del uso que de los términos centrales se hace en diferentes lugares de un mismo marco teórico o el intento de abrir mediante dicho esclarecimiento vías para el desarrollo y perfeccionamiento de teorías y métodos.

Una objeción habitual al programa naturalista en filosofía ha venido consistiendo en indicar que, dado que la tesis epistemológica fundamental del naturalismo implica que no existe una brecha entre los métodos a disposición del filósofo y los métodos a disposición del científico, reduciéndose el escenario de posibilidades a diferentes aplicaciones del método científico, que sería en este contexto entendido como el único medio válido para la obtención de formas fidedignas de conocimiento, y dado que no existe manera de definir dicho método de forma unívoca, esto es, dado que resulta cuando menos comprometido afirmar que existan una serie de principios metodológicos discretos compartidos por las diferentes formas de actividad que caracterizaríamos como científicas, entonces, las dudas acerca de la existencia de algo a lo que quepa denominar "método científico" se extenderían al programa naturalista en su conjunto que, desde esta perspectiva, estaría cimentando castillos en el aire. Esta objeción evidencia una concepción optimista de nuestras prácticas definitorias y su alcance: si algo existe, desde este punto de vista, puede ser definido de forma discreta e inequívoca. Según esta objeción, pues, el hecho de que no quepa ofrecer una definición compacta y unívoca del método científico tendría que hacernos desconfiar de la existencia del mismo y, por tanto, asimismo de las posibilidades del programa naturalista. Cabe rechazar esta objeción de diversas formas, esencialmente de dos: 1) ofreciendo una definición del método científico suficientemente amplia como para dar cabida a prácticas tan heterogéneas como los

estudios longitudinales de medidas repetidas en epidemiología y los de mutagénesis por radiación ultravioleta o agentes químicos, y 2) pidiendo a los objetores que sean ellos quienes ofrezcan una definición unívoca e irreprochable, completa y consistente, por así decir, de «conciencia», «gen» o «filosofía», haciéndoles notar que su fracaso traería consigo nuestras dudas acerca de la existencia de los referentes de esos términos. Ambas formas serían igualmente eficaces, pero la segunda es no sólo más rápida, sino además hacedera.

Otra objeción habitual adopta la siguiente forma: "los filósofos naturalistas han sucumbido, aceptando de forma sumisa y acrítica los estándares científicos". Esta objeción, por su parte, carece de toda fuerza, y ello porque necesita poner patas arriba aquello que critica para hacerlo criticable. ¿Qué es lo que queremos decir con esto? Sencillamente, que no hay nada que aceptar en las ciencias, porque ellas no exigen adherencia, sino más bien todo lo contrario. Como una rápida consulta de cualquier manual de metodología, fundamentos de investigación, diseños experimentales o análisis de datos evidencia, los estándares científicos que el naturalista, según esta objeción, acepta acríticamente consisten, precisamente, en no aceptar nunca nada de forma definitiva y en criticar sin miramiento cuanto no haya sido aún rechazado. Así, por ejemplo, las hipótesis científicas no se aceptan sino provisionalmente mientras sigue buscándose evidencia en su contra. Imputar sumisión al naturalista es, pues, un contrasentido, porque si existe una actividad humana a la que el apelativo «crítica» le venga como un guante no es otra que ésa a la que viene denominándose "científica". El naturalismo no tiene nada que aceptar sumisamente, porque consiste, justamente, en un intento de rechazar.

### 3._Filosofía de la mente naturalista

Por lo que toca al extremo que nos ocupa, la principal consecuencia que cabe extraer de lo hasta aquí apuntado es la de que un filósofo de la mente interesado en contribuir positivamente en la labor interdisciplinaria de las ciencias cognitivas debiera procurarse una sólida formación en al menos alguna de las restantes disciplinas que las integran, y ello, entre otras cosas, porque, tal y como queda ya implícito en lo antedicho, "la característica central (…) de la filosofía actual de la mente es que ella misma está integrada en el campo de las ciencias cognitivas" (Martínez-Freire, 1996: 302). A este respecto, Daniel Dennett, en el prólogo a la segunda edición de su primer libro –su tesis doctoral (Dennett, 1969/1986/2010)–, anota entre los aciertos del mismo el enfoque

naturalista que articula sus planteamientos, y subraya que en la época en la que lo redactaba no resultaba habitual que los filósofos de la mente se preocuparan de mantenerse informados en ciencias cognitivas, mientras que ahora –Dennett escribe este prólogo en 1985– las cosas han cambiado: "A fairly professional knowledge of the other cognitive sciences –psychology, artificial intelligence, linguistics– is now considered a virtual qualification for professional status in the discipline" (Dennett, 1969/1986/2010: xv del original, 11 de la traducción).[10] En esta línea, cabe apuntar a la existencia de dos condiciones impuestas a la participación del filósofo en ciencias cognitivas: "suficiente conocimiento de los campos disciplinarios (…) relevantes en el estudio de su objeto específico de interés y capacidad [para tender puentes entre] dichos campos" (González, 2009: 69).

La filosofía de la mente ha experimentado desde la década de los cincuenta una evolución que podemos seguir sin apenas salirnos del ámbito de la así llamada tradición analítica. Dentro de esta tradición han proliferado tanto los conatos naturalistas como las críticas, las cautelas y el escepticismo hacia las posibilidades del proyecto de naturalizar la mente ofreciendo las claves para una comprensión científica de la misma. Entre los conatos naturalistas podríamos, con cierta liberalidad, incluir a las teorías de la identidad, los funcionalismos y la neurofilosofía, mientras que entre las referidas críticas y cautelas encontraríamos las más variadas propuestas, centradas habitualmente en argumentos ideados para atacar los supuestos en que se basan los primeros.[11] Podríamos aludir, para ilustrar breve y parcialmente el cariz de este segundo conjunto de planteamientos, a la plétora de experimentos mentales e intuiciones misterianas que pueblan la literatura contemporánea, a la tesis de la superveniencia desarrollada por Davidson[12] y

[10] En los últimos compases de *A Brief History of Analytic Philosophy: From Russell to Rawls*, Stephen P. Schwartz ha extendido esta idea al trabajo filosófico en general, pronosticando una creciente interdisciplinariedad y destacando que una sólida formación en las areas pertinentes de las ciencia sociales y naturales resulta ya a día de hoy imprescindible para cualquiera que pretenda hacer contribuciones significativas en filosofía de la mente o filosofía del lenguaje (Schwartz, 2012: 321).

[11] Estos ataques y críticas no han dejado de adoptar las más acerbas de las formas, llegando a ser en ellas calificadas determinadas formas de naturalismo de "descerebrado cientifismo cerebralista" –así en Rodríguez González (2006: 330), recensión, ingeniosamente intitulada con reminiscencias sartreanas, del manual de Carlos J. Moya (Moya, 2004).

[12] Nuestra inclusión de Davidson en este grupo obedece a la imposibilidad de integrar lo mental en un contexto nomológico de explicación natural a que conduce su monismo anómalo –imposibilidad que se hace patente, por ejemplo, en la primera frase o en el famoso tercer principio de "Mental events" (Davidson, 1970: 207, 208, respectivamente). Somos conscientes, por otra parte, de que el marco general del planteamiento de Davidson es habitualmente tratado *como si* de una propuesta naturalista se tratara (él mismo, en "Epistemology externalized", pretendió que así debía entenderse): su epistemología de tercera

más recientemente por Chalmers –con la cual asistiríamos al *naturalismo mínimo* de un materialismo no reductivo que, en casos extremos, como el de Chalmers, habría llegado a desembocar en un panprotopsiquismo dualista–, al subjetivismo de Nagel o al epifenomenalismo de Jackson. Entre ambos bloques, aunque más cerca del segundo, entendemos, cabría ubicar el naturalismo biológico de Searle o el cartesianismo naturalizado de Galen Strawson, posturas ambas pretendidamente conciliadoras pero asentadas sobre la sonora aunque exigua base de la iteración de una tesis según la cual carece de sentido definir lo físico por contraposición a lo mental.

Pero, ¿qué cabe meter en el saco de las propuestas naturalistas? ¿Qué incluimos en el conjunto de las mismas? Una primera aproximación puede entresacarse ya de lo expuesto a lo largo del presente capítulo: en este saco caerían aquellas contribuciones en las que los planteamientos filosóficos y los actuales desarrollos científicos, tanto empíricos como teóricos, se hallan estrechamente relacionados, adoptando esta relación una forma bidireccional: el filósofo naturalista puede proponerse tanto desarrollar una ontología de lo mental alineada con los desarrollos de las disciplinas encargadas del estudio científico de la mente, el cerebro y la conducta como abrir camino a esas disciplinas mediante el esclarecimiento conceptual y la crítica de marcos teóricos y programas metodológicos. Una caracterización general del contenido de este cajón de sastre del naturalismo en filosofía de la mente habría de atender, pues, fundamentalmente a dos extremos. En primer lugar, debiera aludir al naturalismo ontológico, que en su versión estricta vendría a defender que los *fenómenos mentales*[13] verdaderamente existentes y, por tanto, aquéllos de los que cabe hablar con sentido son fenómenos enteramente susceptibles de un análisis empírico y no serían, de este modo, otros que los fenómenos que a las ciencias de la mente, el cerebro y la conducta les cabe estudiar. El problema consistiría en este punto en la elección de disciplina y paradigma, dado que una efervescente pluralidad caracteriza cada disciplina y, además, en el seno de cada una de ellas –y aun en el de determinados paradigmas dentro de cada una de ellas– se utilizan los mismos conceptos en referencia a fenómenos que, aunque puedan ser tenidos por análogos, e incluso idénticos, son designados de forma diversa en función de los distintos marcos conceptuales en que se encuadren. No obstante, a pesar de este problema, la asunción

persona, su ontología monista... que, sin embargo, desembocan en un dualismo metodológico que ubica a *lo humano* más allá del alcance del potencial explicativo de las ciencias naturales.

[13] «Fenómeno mental» es la locución aceptada. Cuando el filósofo de la mente defensor de una ontología naturalista utiliza la locución «entidad mental», lo hace para negar la existencia de "entidades" mentales, en línea con autores como Paul y Patricia Churchland, Paul Feyerabend, Richard Rorty o Stephen Stich, todos ellos, a su vez, en la línea del quineano abandono del lenguaje intencional.

general según la cual desde donde nos cabe iluminar la naturaleza de los fenómenos mentales es desde dentro de los márgenes del método científico no resulta opcional para el naturalista en filosofía de la mente y, así, sus supuestos ontológicos difícilmente pueden enunciarse de forma independiente de sus supuestos epistemológicos: los únicos rasgos que compartirían todos los fenómenos mentales efectivamente existentes serían para el naturalista los siguientes: a) que, de ser desentrañables, habrán de serlo gracias al método científico y b) que a ninguno de ellos le es en ningún sentido dable contravenir las leyes fundamentales de la naturaleza (resultando de especial interés las leyes de conservación, particularmente el primer principio de la termodinámica). En segundo lugar, una tal caracterización debiera atender al plano metodológico ofreciendo orientación a la hora de arrostrar la cuestión del papel que al filósofo le cabe desempeñar en el estudio de la mente cuando se asume que lo mental ha de ser analizado haciendo uso del método científico. La intervención del filósofo en el estudio de lo mental puede adoptar, desde luego, muy diversas formas. Nada le impide vestirse de metodólogo y hacer filosofía de la ciencia aplicada, o de crítico sociocultural y ofrecer perspectivas acerca del modo en que determinadas investigaciones, prácticas o resultados científicos encajan en el marco general de nuestra cultura contemporánea. Pero puede también el filósofo navegar en el barco de Naurath implicándose positivamente en el trabajo de la comunidad investigadora interesada en allanar y prolongar el camino hacia un conocimiento seguro de la naturaleza de lo mental. Un filósofo inclinado a alguna de las señaladas actividades –metodología aplicada, crítica sociocultural– puede perfectamente tenerse por naturalista, pero los quehaceres propiamente naturalistas tenderían antes bien a ubicarse en la estela de la labor del referido navegante, una labor que sazona no sólo en forma de estudio analítico de los límites, oportunidades e implicaciones de las teorías posibles acerca de lo mental (Dennett, 1988: 275), o de informada crítica y contribución al esclarecimiento de los fundamentos de las diversas disciplinas cognitivas y del sentido y funcionalidad tanto de las nociones más básicas como de las más generales dentro de las mismas, sino asimismo en forma de aportaciones teóricas encarnadas en hipótesis, herramientas lógicas, conceptuales, metodológicas y heurísticas. En cualquier caso, los dos extremos de nuestra sucinta caracterización del naturalismo en filosofía de la mente son en buena medida independientes y, de este modo, puede el filósofo naturalista contribuir en este segundo plano metodológico sin preocuparse del ontológico –que es precisamente lo que nos proponemos hacer a lo largo del resto de esta segunda parte.

*Delineado ya el marco naturalista de nuestra discusión de los medios adecuados para la explicación de la conciencia –en otras palabras, para la solución del problema de la conciencia–, comencemos por examinar desde el mismo aquéllos que, a pesar de su inadecuación, han predominado en el estudio y en el debate contemporáneo acerca de la conciencia. De entre ellos, prestaremos atención prioritaria a la idea de que una tal explicación depende de la especulación a priori acerca del significado y la interrelación entre las nociones de representación y conciencia. Por extraño que parezca, ésta es la actividad a la que viene dedicándose la mayoría de los filósofos implicados en los* Consciousness Studies. *Los motivos de esta tendencia hay que buscarlos en la ya prolongada confrontación en ciencias cognitivas entre partidarios y detractores de una metafísica de lo mental cortada por el patrón del reduccionismo representacional. Pero no nos encargaremos de las raíces u orígenes de dicha tendencia, sino más bien de sus consecuencias y, decisivamente, de poner de relieve el escaso provecho que a la investigación de la naturaleza de la conciencia le cabe extraer de esta clase de lides abstractas.*

# CAPÍTULO 8

## LA CUESTIONABLE FERACIDAD DE LOS TÉRMINOS QUE ARTICULAN EL DEBATE CONTEMPORÁNEO

### 1._Naturalismo, mente y conciencia

Según Searle (1992: 2 del original, 16 de la traducción), los filósofos de la mente materialistas entienden que "naturalizar" un fenómeno metal consiste en "reducirlo" a un fenómeno físico. Con esto pretende poner de relieve la desorientación de tales filósofos, pues con «reducir» quiere significar algo así como "negar que cualesquiera otros rasgos o propiedades de tales estados, más allá de su efectivo consistir en dinámicas de puntos de masa/energía, no son sino fictivos delirios amalgamados con prejuicios tradicionales". Es decir, en su interpretación, naturalismo –excepto el suyo, el biológico– equivale siempre a materialismo, reduccionismo y, en último término, eliminativismo. De este modo, cuando Hans-Johan Glok afirma que al enfrentarse con aparentes contraejemplos el naturalista tiene dos opciones –"she can either dismiss them as spurious or try to show that on closer scrutiny they boil down to a scientific or natural phenomenon" (Glock, 2008: 139)–, Searle defendería que en el caso de la conciencia cuenta, de hecho, sólo con una –la primera.

La cuestión de la naturalización de la mente fenoménica ha sido planteada desde que de la misma comenzaran a ocuparse diferentes especialistas en ciencias cognitivas en unos términos, entendemos –y argumentaremos–, equívocos. La mente, ha venido arguyéndose, se caracteriza de forma exclusiva y exhaustiva por su capacidad representacional (esto es, por hallarse dotada de intencionalidad) y su aspecto experiencial o

cualitativo (esto es, por hallarse infundida de fenomenalidad). Así, ha venido asumiéndose que la mente consta de dos "elementos", designados por aquellos descriptores ("capacidad representacional", "aspecto experiencial"), y, adicionalmente, que diversos paradigmas en ciencias cognitivas podrían terminar dando acabada cuenta del primero. Nadie sabe de qué manera cabría probar que cosa semejante pudiera o no suceder o haber sucedido ya –y de hecho los paradigmas computacionales y las nuevas tendencias situadas y corpóreas o encarnadas parece que seguirán conviviendo y explicando los mismos o diferentes procesos cognitivos en referencia a distintos mecanismos y con arreglo a marcos conceptuales tan diversos que en algunos de ellos la propia noción de representación no juega un papel definido–, pero la asunción habitual consiste en pretender que la mente intencional no supone mayor problema para el naturalista, configurándose de este modo dos bandos: el de los que, con Dennett, Dretske, Harman, Lycan o Tye, han venido argumentando que la mente fenoménica puede reducirse a la representacional, la cual, a su vez, habrá de un modo u otro de mostrarse dócil a nuestro acceso epistémico, y el de los que, con Block, Chalmers, Loar, Searle o Peacocke, han venido defendiendo la autonomía e irreductibilidad del ámbito de lo mental que lo fenoménico representaría.

Ha venido considerándose, pues, que dado que lo fenoménico es un fenómeno natural que no sabemos muy bien de qué modo cabría integrar en un marco explicativo naturalista, sólo nos caben dos opciones: tratar de probar a) que puede reducirse a otro hipotéticamente concreto, discreto y aproblemático (la intencionalidad), por más que los paradigmas empíricos destinados a dar cuenta del mismo se encuentren actualmente en constante liza, o b) que, sencillamente, es autónomo e irreductible y, de hecho, más básico que lo intencional. La primera opción parece un mero acto de fe: no disponiendo de una sólida base en que apoyar el habitual supuesto según el cual la mente intencional puede encuadrarse en paradigmas teórico-experimentales perfectamente legítimos con mayor facilidad que la fenoménica, no acaba de presentarse como una opción enteramente cabal e indigna de algunos más que sanos escrúpulos la de comenzar a construir la casa por el tejado elaborando especulativos marcos conceptuales dentro de los cuales resulte lo fenoménico, supuestamente, explicable en términos intencionales, y lo intencional, a su vez, en virtud de un a día de hoy indeterminado programa teórico-metodológico capaz de dar acabada cuenta del modo en que arraiga en procesos biológicos completamente ordinarios. Así las cosas, parece que la segunda opción habría de presentársenos como más atractiva, pero entendemos que se basa en asunciones tan in-

fundadas como la primera. Los autores que rechazan la naturalización de la fenomenalidad de la mente a través de su reducción a la mente intencional o bien sostienen que la mente fenoménica no puede explicarse científicamente y abrazan el misterianismo, o bien que no puede reducirse a la mente intencional porque es más básica que ella y, por tanto, para poder hablar de auténticas representaciones mentales éstas han de estar acompañadas de conciencia fenoménica. Argumentaremos, por nuestra parte, que esta estrategia *inseparatista* no hace sino complicar el ya de por sí complicado debate en torno al problema de la conciencia.

Comenzaremos por ofrecer algunas pinceladas acerca de la noción de representación, criticando por el camino la endémica presuposición según la cual «intencionalidad» y «fenomenalidad» son etiquetas léxicas unívocas, homogéneas y aproblemáticas, para pasar a continuación a criticar tanto la solución separatista –esto es, el reduccionismo representacional: el intento de naturalizar la mente fenoménica reduciéndola a la representacional– como la inseparatista –esto es, el antirreduccionismo fenoménico, derivado en una suerte de fundamentalismo fenoménico según el cual sin experiencia no hay representación.

## 2._La oscura claridad de los términos que articulan el debate contemporáneo

Para comenzar a delimitar el sentido en que suele emplearse la noción de representación en ciencias cognitivas conviene echar mano de la distinción entre vehículos representacionales y contenidos representacionales. Según la misma, tendríamos, por una parte, aquellos objetos o eventos que representan o en los que la representación tiene lugar (los vehículos) y, por otra, aquellos objetos o eventos que son representados (por los vehículos) como siendo de uno u otro modo (vid. Dretske, 1995: 34 y ss.). Haciendo uso de esta distinción cabe considerar cualquier vehículo representacional como un proceso mental biológicamente implementado y su contenido como su significado y su capacidad para influir en la conducta del organismo que sea el caso. No obstante la utilidad de esta distinción, la misma no debe hacer que perdamos de vista el modo en que constituye un ejercicio de abstracción consistente en separar dos conceptos entre cuyos referentes, en la práctica, quizá resulte más que problemático trazar fronteras. Con todo, lo que nos interesa aquí es subrayar que para que quepa hablar de representación es necesario contar con un organismo en el cual determinados conjuntos de procesos (los vehículos) cumplan la función de referir a determinados eventos claramente diferencia-

dos de aquellos conjuntos (los contenidos). Plantear la pregunta acerca del sentido en que determinadas formas de relación entre un organismo y su medio permiten hablar de vehículos y contenidos representacionales trae consigo las dos siguientes cuestiones. En primer lugar, la cuestión acerca de la clase de relación que se da entre una representación y aquello que representa. En segundo lugar, la cuestión de la propia naturaleza de las representaciones. Un resoluto vistazo a la historia de las ciencias cognitivas obliga a admitir que los conatos hasta la fecha incoados con vistas a abordar ambas cuestiones han sido en cualquier caso insuficientes y no han alcanzado a ofrecer una operacionalización adecuada y consensuada de la noción de representación, motivo por el cual a nadie extraña que se alcen voces afirmando que el proyecto de naturalizar dicha noción es un desatino basado en un error conceptual (Horgan, 1992; 1994), que la misma es enteramente innecesaria (Brooks, 1991) o sistemáticamente mal utilizada en áreas de investigación como las neurociencias (Ramsey, 2007).

Las respuestas tradicionales al problema de la naturaleza de las representaciones –al que Cummins (1989: 1) se refiere como el problema de las representaciones (en plural)–, las encontramos en las diferentes interpretaciones de la concepción computacional de la mente compartida por simbolistas y conexionistas. Tanto para unos como para otros la mente puede ser caracterizada como un sistema que procesa información. El procesamiento de información tiene un obvio carácter representacional: la información que manipula un sistema computacional debe ser representada de alguna manera, existiendo una relación entre el modo en que tienen lugar las operaciones de cómputo y el modo en que la información es representada por el sistema en que dichas operaciones se realizan. A pesar de diferir radicalmente en su forma de dar cuenta del carácter representacional de la mente, simbolistas y conexionistas comparten una perspectiva internalista al tratar de explicar la conducta o actividad cognitiva de los organismos mediante la referencia a procesos internos o sistemas de estados internos. Según Martínez-Freire (1992), esta perspectiva internalista, sumada a la concepción del sujeto como procesador de información, sirven para definir el núcleo de la concepción cognitivista de la mente. Dentro de esta concepción cabe distinguir, como sugeríamos, entre la interpretación de la noción de representación que ofrece el marco teórico del simbolismo clásico y la que ofrece el del conexionismo. La interpretación simbolista presenta la actividad interna de carácter representacional postulada por el cognitivismo en términos de procesamiento de información en forma de símbolos, y a los símbolos como físicamente implementados, discretos, de una pieza, determinantes para el pergeño de la conducta y

semánticamente interpretables. Desde el punto de vista de esta interpretación, esa actividad representacional consiste en una serie de procesos computacionales basados en la manipulación de símbolos discretos de acuerdo con una sintaxis combinatoria. En el marco simbolista la conducta cognitiva es explicada recurriendo a procedimientos de cálculo recursivos análogos a los utilizados en el cálculo aritmético. Estos procedimientos operarían con los señalados símbolos mediante la aplicación de algoritmos previamente establecidos, de forma y manera que el organismo alcance a ofrecer la respuesta conductual adecuada en virtud de los insumos perceptivos disponibles y la estructura del sistema que integrarían los símbolos codificados y las reglas almacenadas para manipularlos. La gran cantidad de procesos mentales difícilmente abordables desde dentro del marco simbolista sirvió para incitar a los investigados a buscar en nuevas direcciones, propiciando el origen de un nuevo paradigma en ciencias cognitivas: el conexionismo. Las señaladas carencias del paradigma simbolista se hallaban crucialmente relacionadas con los niveles más bajos en la jerarquía del procesamiento de la información, como la percepción, pero también con los procesos cognitivos de alto nivel, como el razonamiento, en los que se entendía que el paradigma simbolista debería desempeñarse con total solvencia –cabe aludir en este sentido al comentadísimo "frame problem" (McCarthy & Hayes, 1969)–. Con la emergencia del conexionismo, un paradigma que pronto se mostraría capaz de modelar con éxito tareas inabordables dentro del marco simbolista (como el reconocimiento de patrones lingüísticos o visuales), la noción cognitivista de representación vira de la discreción a la dispersión: en las redes conexionistas la representación se produce no en elementos sino entre elementos. Se trata de representaciones distribuidas en redes que remedan grupos neuronales naturales, redes cuyos transitorios subconjuntos de estados globales constituirían los vehículos representacionales. De este modo, en la interpretación conexionista, las representaciones no serían entidades discretas, de una pieza, no se hallarían puntualmente localizadas (nada de *grandmother cells*), sino que dependerían de propiedades sistémicas. Simbolismo y conexionismo comparten su carácter computacional –sus explicaciones de la conducta y la actividad cognitiva se alzan por igual sobre la idea de cómputo mental–, pero los partidarios del conexionismo rechazarían la idea de un sistema simbólico operable en virtud de un conjunto de reglas. La actividad representacional deja de atribuirse a símbolos codificados, a programas almacenados o a lenguajes internos entendidos como sistemas de reglas para la manipulación de símbolos y pasa concebirse como distribuida en la disposición de fuerzas o pesos de las conexiones entre las unidades de una red: en una

red conexionista la información se halla almacenada de forma no simbólica en la distribución de fuerzas de conexión entre unidades. En este sentido, Patricia S. Churchland y Rick Grush definen la noción de representación en sistemas conexionistas como sigue: "Representation is characterized as a pattern of activation across the units in the neural net, where this can be described as a vector, and computations are therefore vector-vector transformations" (Churchland & Grush, 1999: 156). Ya en la década de los ochenta Paul M. Churchland definía las representaciones en el marco de los patrones de actividad neuronal como puntos y trayectorias en el espacio vectorial y las computaciones sobre representaciones como transformaciones de vector a vector. (vid., v. g., Churchland, 1984/1987/2013: 232 y ss.; 1988: 127; 1989: xiii-xiv; 1992; 2005). La perspectiva simbolista y la conexionista coincidirían, pues, en su concepción computacional de la mente, pero desde interpretaciones diversas de la noción de cómputo: "el modelo de computación que constituye el ideal de la psicología cognitiva clásica consiste en la aplicación serial o sucesiva de reglas formales almacenadas de manera explícita sobre representaciones simbólicas explícitas y localizables de manera definida. En cambio, el modelo de computación que constituye el ideal de la psicología cognitiva conexionista consiste en la aplicación en paralelo de reglas no predeterminadas de manera fija (sino flexibles) sobre unidades de activación conexionadas que representan al nivel de la red constituida por las unidades y no al nivel de cada unidad localizable" (Martínez-Freire, 1995: 148). No obstante, a pesar de ofrecer respuesta –mediante el modelando de diversas formas de conducta o actividad cognitiva de bajo nivel– a problemas que el paradigma simbolista no podía afrontar, a menudo se ha señalado que el paradigma conexionista no se muestra capaz de explicar las capacidades mentales de alto nivel, aduciendo los partidarios del simbolismo que dentro del paradigma conexionista la sistematicidad y la productividad propias de las capacidades cognitivas de alto nivel, como el uso del lenguaje, de ningún modo pueden hallar acomodo (Fodor & Pylyshyn, 1988; Fodor & McLaughlin, 1990).

En cualquier caso, entendamos que la naturaleza de la representación es la de los símbolos discretos manipulables mediante reglas fijas aplicadas en serie o bien entendamos que la misma corresponde con la de vectores de activación manipulables mediante reglas flexibles aplicadas en paralelo, nos queda por responder a la pregunta acerca de la clase de relación que se da entre una representación y aquello de lo que la misma es una representación. Ésta es la cuestión que Cummins (1989: 2) ha denominado el problema de la representación (en singular). De si cabe o no ofrecer una solución

aceptable a dicho problema depende que pueda disponerse de una noción satisfactoria de representación en ciencias cognitivas e, incluso, la posibilidad de operacionalizar de forma válida dicha noción.

Los conatos ensayados en esa dirección se han presentado, esencialmente, en las siguientes variedades: las teorías causales, las teleosemánticas y las del rol funcional. Las teorías causales –denominadas también informacionales a causa del papel decisivo que la teoría matemática de la información jugara en la elaboración de una de las más tempranas e influyentes de estas teorías, la de Dretske (1981; 1983)–, defienden que la naturaleza de la relación representacional es de tipo causal, de tal modo que mi instancia representacional "estrella" es acerca de las estrellas en tanto son ellas las que producen en mi economía cognitiva un determinado designar una instancia representacional "estrella" concreta. Así, mi instancia representacional "Sol" representa al Sol porque es causada por éste. Una teoría estrictamente causal sería, no obstante, incapaz de dar respuesta al denominado problema de la disyunción (Fodor, 1984). Este problema se presenta para las teorías causales desde el momento en que un puñetazo en un ojo produce en mí un determinado designar una instancia representacional "estrella" concreta. ¿Por qué? Porque en esta situación una teoría estrictamente causal de la representación debería admitir que la representación "estrella" tiene como contenido la disyunción *estrella-o-puñetazo-en-un-ojo* en lugar del contenido no disyuntivo *estrella*, dado que tanto las propias estrellas como un puñetazo en un ojo causan instancias de dicha representación. Dentro de las teorías causales se han propuesto diferentes soluciones al problema de la disyunción. El propio Fodor trató de hacer frente al mismo con su teoría de la dependencia causal asimétrica (vid. Fodor, 1987; 1990a; 1990b; 1994), en cuyo núcleo se halla la idea de que la causa determinante del contenido representacional (las estrellas, en nuestro ejemplo) y la causa no determinante del mismo (un puñetazo en un ojo, en nuestro ejemplo) tienen diversos grados de relevancia dado que la segunda ha de basarse en la primera, que sería desde este punto de vista la fundamental: los puñetazos en un ojo, o el LSD, no podrían causar esa representación de no existir la causa determinante.

Otros autores, como Ruth Millikan, han involucrado a la teoría de la evolución en su elaboración de una clase diversa de teorías de la representación capaz de hacer frente a este problema, una clase de teorías que en ocasiones se encuentra clasificada como funcional y en ocasiones como causal: la teleosemántica. La naturaleza de la relación representacional, desde el punto de vista teleosemántico, incluye la historia filogenética del organismo: la relación representacional tiene la forma que tiene dada la historia evo-

lutiva de la especie que sea el caso y cumple una función determinada en virtud de la misma. Sin entrar en detalles, desde la perspectiva teleosemántica de Millikan, la más exitosa y discutida, una característica K de un organismo contará como una representación si y sólo si: a) se ubica entre los segmentos productores y consumidores de representaciones del organismo, seleccionados para ajustarse mutuamente, b) si dicha característica tiene como función propia (proper function) adaptar al consumidor de representaciones pertinente a determinado aspecto del entorno, permitiendo al organismo conducirse adecuadamente con respecto a dicho aspecto, y c) si existe una serie de transformaciones de K cuya función consista en adaptar a los consumidores de representaciones pertinentes a las transformaciones correspondientes del aspecto relevante del entorno.

No obstante, tanto la solución de Fodor como la teleosemántica han sido objeto de sucesivas críticas a las que resulta más que comprometido afirmar que hayan logrado ofrecer contrarreplicas eficaces. Entre las más enjundiosas de estas críticas, entendemos, se encuentran aquéllas que acusan a Fodor de violar el punto de partida naturalista de las teorías causales al postular una asimetría causal cuyo mecanismo, en cualquier caso, habría que explicar en términos naturalistas, lo cual parece hallarse fuera de las posibilidades de la teoría de Fodor (vid. Seager, 1993; Adams & Aizawa, 1994; 1994b; Gibson, 1996), y aquéllas que, frente a las apelaciones a la teoría de la evolución, ponen de manifiesto las lagunas presentes en la interpretación que estas propuestas hacen de dicha teoría y de la noción de función biológica (vid. Blackmon, Byrd, Cummins, Lee & Roth, 2006; Jaume Rodríguez, 2012: 133-135; Fodor, 1990a; 1990c).

Por su parte, las teorías del rol funcional postulan que una representación es, efectivamente, una representación de algo en función del rol que la misma desempeña en la economía cognitiva de un organismo. La relación representacional deja así de ubicarse entre la representación y aquello que ella representa para ampliarse al incluir segmentos mayores del aparato y la dinámica cognitiva del organismo: se trata de una clase de relación que, en paralelo a la filosofía del lenguaje del segundo Wittgenstein, incluiría en este marco el uso de la representación que sea el caso, así como sus interacciones –también potenciales– con otras representaciones disponibles, una clase de relación que no contemplaría exclusivamente las interacciones que pudieran darse dentro del propio sistema cognitivo del organismo, sino asimismo con elementos externos al mismo –y es en este sentido que se habla de teorías bifactoriales (Loar, 1981; Block,1986)–. Lo que determina que una representación sea una representación de aquello que representa es, desde esta perspectiva, el papel que juega en la vida cognitiva del agente en que se ins-

tancia. Un obstáculo que se presenta a este tipo de planteamiento funcionalista consiste en que para definir la naturaleza de la relación representacional necesita –al postular que es la relación con otras representaciones lo que hace de una representación la representación que es–, apelar a la propia noción de representación, que subyacería así como un supuesto difícilmente explicitable dentro de este marco teórico holista. Por otra parte, precisamente la concepción holística de la naturaleza de la relación representacional defendida por las teorías del rol funcional implica que cualquier cambio en cualquier lugar de un sistema representacional dado modificará el carácter de, virtualmente, cualquier relación representacional posible dentro del mismo, provocando que el uso que cualquier organismo pueda hacer de cualquiera de sus representaciones sea elevadamente inestable: cuando digo "estrella" no hago lo mismo que Pepe al hacer lo mismo, ni lo mismo que cuando lo haga dentro de media hora.

El óbice que los intentos hasta la fecha realizados con la intención de abordar el problema de la representación (en singular) se muestran incapaces de soslayar es, sin embargo, el de ofrecer una caracterización de la noción de representación que no sea aplicable incluso a una bicicleta (Haselager, De Groot, & Van Rappard, 2003: 17), es decir, el de establecer criterios que sirvan para delimitar la aplicación de la noción de representación e identificar así aquellos organismos o sistemas que usan representaciones diferenciándolos de los que no lo hacen.[1] La respuesta a este impasse ha adoptado dos formas, consistiendo bien en negar –habitualmente desde perspectivas dinamicistas

[1] Este punto suele ilustrarse mediante ejemplos de dispositivos que cuesta creer que manipulen representaciones pero que pueden concebirse sin dificultades como cumpliendo los criterios planteados por las diferentes teorías de la representación. El regulador centrífugo de Watt-Boulton es el ejemplo habitual. Así, por ejemplo, van Gelder ha descrito al regulador centrífugo como un dispositivo computacional que implementa un algoritmo –que asimismo detalla el autor (Van Gelder, 1995: 348)– para el cálculo de los cambios pertinentes en el ángulo de la válvula reguladora. Por su parte, Bechtel (1998) ha mostrado que este dispositivo cumple los requisitos planteados por las interpretaciones funcionales de la noción de representación. De forma escueta, estos criterios incluirían la presencia de: 1) un objeto o evento X que es representado por el sistema; 2) un objeto o evento Y que representa el objeto o evento representado; y 3) un sistema Z que utiliza Y para coordinar su comportamiento con X. Todos estos elementos están, según Bechtel, presentes en el regulador centrífugo de Watt-Boulton: X sería la velocidad del volante de inercia, Y el ángulo del brazo del regulador, que representa la señalada velocidad, y Z el mecanismo de la válvula que utiliza Y (abriendo o cerrando la válvula) para coordinar el comportamiento del dispositivo con X. La interpretación teleosemántica de la noción de representación tampoco se libra de esta criba. De forma sucinta, el regulador centrífugo de Watt-Boulton podría ser concebido como un sistema representacional en términos teleosemánticos como sigue: el segmento productor de representaciones (el mecanismo del volante de inercia) y el segmento consumidor (el mecanismo de la válvula) han sido diseñados para encajar entre sí y ajustarse mutuamente y el brazo giratorio se halla ubicado entre ambos (requisito a); el ángulo formado por los brazos giratorios tiene como función propia adaptar al consumidor de representaciones (el mecanismo de la válvula) a algún aspecto del entorno (la velocidad del motor), permitiendo al regulador centrífugo comportarse apropiadamente (manteniéndola constante) respecto de dicho aspecto (requisito b); para terminar, existe una serie de transformaciones del ángulo de los brazos cuya función consiste en adaptar al mecanismo de la válvula a las transformaciones correspondientes de la velocidad del motor (requisito c).

como las gestadas a partir de las intuiciones de van Gelder (1995)– la necesidad en que las ciencias cognitivas se verían de recurrir a la noción de representación, bien en tratar de restringir dicha noción de tal modo que pueda evitarse la ubicuidad de su aplicabilidad reservando la etiqueta "representacional" exclusivamente para procesos acaecidos en aquellos organismos capaces de desarrollar procesos cognitivos superiores. El problema que surge cuando tratamos de restringir de este modo la noción de representación es que la de procesos cognitivos superiores es igualmente equívoca e implica nuevamente la necesidad de improvisar criterios para delimitar los procesos cognitivos superiores de los inferiores.

La cautela se muestra como la actitud más natural ante perspectivas esencialistas de tipo todo o nada y, por lo tanto, asimismo ante la posibilidad de elucidar criterios para delimitar con claridad lo que es propiamente intencional o representacional respecto de lo que no lo es. Clark & Toribio (1994) argumentaron hace ya dos décadas que desechar el proyecto de destilar tales criterios podría resultar más fructífero que continuar ahondando en la tarea de repentizarlos para ofrecerlos a su inmediata confutación. Posteriormente, Clark & Grush (1999) extendieron este planteamiento a los criterios para distinguir entre sistemas cognitivos y sistemas no cognitivos. Tanto en un caso como en el otro, el tiempo no les ha quitado la razón y tales criterios siguen sin aparecer por ninguna parte. La de representación parece ser, en cualquier caso, una noción singular que viene pretendiendo predicarse de una realidad plural y, por añadidura, con fronteras filogenéticas tanto o más difusas que las de la de conciencia fenoménica, unas fronteras dentro de las cuales bien cabe que quepa distinguir grados, diferentes formas y niveles de representación, como bien cabe que haya de hecho saltos bruscos entre esos grados, pero también que nos sea imposible estipular criterios excesivamente estrictos para delimitarlos, y muy probablemente nos encontremos ante un complejo y plural continuo en el que no les sea a nuestras categorías dable hacer acabadamente justicia a sus objetos, porque "no hay rigor terminológico posible donde no haya fronteras marcadas" (Morin, 1987: 64 de la traducción). Así, tratar de restringir la aplicabilidad de la noción de representación mediante criterios rigurosamente definidos puede resultar una tarea entretenida, pero puede asimismo que a la misma se le abran pocas posibilidades al margen de la fungir como pasatiempos para lexicógrafos. Proust (1999), por ejemplo, reserva su noción de representación a aves, reptiles y mamíferos, mientras otros la extienden a todos los vertebrados, otros a los insectos, otros a cualquier ser vivo y aun otros a las partículas elementales. Hay, en definitiva, entre la mínima *actividad* percep-

tiva (Noë, 2004) concebible y la capacidad para representar los estados representacionales de otros organismos un buen espacio para improvisar definiciones mínimas o máximas de «representación», para atribuir capacidad representacional exclusivamente a primates o incluso a virus, al igual que hay más bien pocos motivos para suponer que todos y cada uno de estos tipos de actividad biológica habrán de plegarse a los mismos conceptos, basarse en idénticas clases de procesos, reducirse a un discreto puñado de principios abstractos, explicarse partiendo de ellos o provenir de idénticos ríos filogenéticos. Plantea esta última afirmación una enorme serie de extremos que no podemos abordar de forma minuciosa, pero en los que entendemos necesario abundar, por más que sea de forma concisa.

Actualmente se debate fatigosamente acerca del modo correcto de naturalizar la intencionalidad y acerca de si el modo correcto de naturalizar la mente fenoménica es o no vía su reducción a la mente intencional o representacional. Se trata de un debate que suele plantearse en términos de orden explicativo, esto es, en términos de qué explica qué dentro de la esfera de lo mental, y lo que apenas nunca se plantea de forma explícita es que si esto es así y una de ambas instancias de lo mental es más básica y ha de ser por tanto aquélla en términos de la cual sea explicada la otra, quizá ello deba integrarse en un marco teórico en el que razones filogenéticas den cuenta de los motivos por los cuales se supone que uno u otro de estos fenómenos biológicos supuestamente unívocos y monolíticos ha ser explicado en términos del otro. Pero si uno de ellos ha de explicar al otro, al menos, deberíamos poder estar seguros de que nuestras categorías apuntan a clases específicas de fenómenos biológicos discretos y homogéneos, esto es, seguros de que ése en cuyos términos pretendemos explicar el otro sea, a su vez, explicable en unos y los mismos términos. Pero, ¿hemos de suponer que los procesos representacionales –si es que hemos de hablar así– implicados en la navegación por el entorno de las abejas (vid., v. g., Gould, 1990; Gould & Gould, 1988/1995; Menzel, 2009; Menzel et al., 2005) son idénticos que los puestos en juego por la teoría de la mente de un sujeto no autista ante el test Sally-Anne de Simon Baron-Cohen, Alan M. Leslie y Uta Frith? ¿O debemos suponer que una teoría fundamental unificadora habrá de integrarlos en su seno mostrando el modo en que ambos se asientan en una misma base? Si este controvertido extremo no es objeto de controversia, a nadie debiera extrañar que tampoco lo sea el de la univocidad y monoliticidad de nuestras nociones de lo representacional y lo fenoménico, que en rarísimas ocasiones se pone en cuestión. Sin motivo, esto es, sin evidencia empírica ni demostración racional de ninguna clase, ha venido suponiéndose

que las de "mente intencional" y "mente fenoménica" son nociones completamente evidentes que refieren a fenómenos netamente discretos y homogéneos, los cuales, por lo tanto, deberán explicarse dentro de *un* marco teórico análogo. ¿Ha de ser esto necesariamente cierto? No. ¿Por qué? Veamos. Ciertamente, todas las clases de reproducción –del mismo modo que todas las clases de motilidad o todas las clases de metabolismo– comparten algunos principios básicos en todo lo ancho y largo de los cinco –o veintiséis (vid. Drozdov, 2003), tanto da– reinos, pero no dejan de ser principios que no podrían sino compartir por el mero hecho de pertenecer al conjunto de los fenómenos biológicos acaecidos en la tercera roca desde el Sol y no principios que hayan de articular específicamente los diferentes marcos teóricos desde los que cabría abordar la explicación de dicha clase de fenómenos. Esto es, la gemación y la reproducción sexual mamífera pertenecen, en última instancia, al conjunto de los fenómenos naturales que cabe integrar en los marcos teóricos de la biología evolucionista y la genética molecular, pero a nadie se le escapa que de ningún modo cabe extraer conclusiones especialmente interesantes acerca de, por ejemplo, la embriología humana –no hablemos ya de elevar el edificio explicativo completo de la misma– haciendo uso exclusivo de las nociones necesarias para describir y predecir íntegramente el señalado tipo de reproducción asexual. Así, suponer que fenómenos biológicos de mayúsculas dimensiones y que, por añadidura, han recorrido diversos –aunque eventualmente solapados– ríos filogenéticos, emergiendo de ellos, en algunos casos, en sucesivas ocasiones, como la reproducción, la inmunidad, la motilidad, el metabolismo, la mente o la representación son bloques ontológicos macizos que cabe reducir a singulares conjuntos de principios explicativos omnímodos no deja de sonar a ocurrencia, precisamente, un tanto singular, una ocurrencia omnipresente en la bibliografía contemporánea que nadie se ha encargado de justificar. Se busca o se critica *una* teoría de *la* intencionalidad a la cual reducir *la* mente fenoménica. La forma que el vínculo entre este defendido o criticado orden explicativo y el orden filogenético o constitutivo deba adoptar no se discute porque apenas se plantea. Los únicos motivos para defender uno u otro orden explicativo son, en cada caso, exclusivamente ideológicos: no disponemos de sólidos argumentos teóricos o evidencias empíricas en las que apoyar nuestra convicción de que la mente intencional es la base sobre la que ha de alzarse la explicación de la fenoménica –o viceversa–, del mismo modo en que no disponemos de argumentos ni de evidencias en las que basar nuestra convicción de que la mente intencional ha evolucionado a partir de la fenoménica –o viceversa–, del mismo modo en que, finalmente, no disponemos de argumentos ni de evidencias en los que

basar nuestra convicción de que «mente intencional» y «mente fenoménica» son etiquetas aproblemáticas mediante las cuales podemos referir unívocamente a fenómenos homogéneos y discretos explicables desde un único marco teórico fundamental del que, de un modo u otro, habrán de poder derivarse explicaciones particulares de todos y cada uno de los fenómenos que denominamos mentales.

Entendemos, pues, que los términos en los que viene debatiéndose conducen a errores esencialistas y que éstos, sumados a diversas asunciones tradicionales, determinan el modo en que tanto el representacionalismo reductivo separatista como el fundamentalismo fenoménico inseparatista se muestran incapaces de ofrecer ya no respuestas satisfactorias, sino cualquier clase de orientación acerca del modo de abordar el problema de la conciencia.

### 3._Problemas de la solución separatista

El primer problema con que topa la solución separatista es el de la *indeterminación*: nadie sabe cuál es el significado del término «representación», y todos los que creen que lo saben, lo que parecen saber son finalmente cosas muy diferentes. Así, el primer problema se encuentra en la propia noción de representación, dado que, como hemos visto, no cabe hablar hoy día de una noción de representación compartida y ampliamente aceptada en ciencias cognitivas. Dicha noción, lejos de caracterizarse actualmente, pues, por su univocidad, significa cosas diferentes para el partidario del simbolismo clásico, el del conexionismo o el del embodiment y se discute, por otra parte, acerca del formato o código –simbólico, lingüístico, pictórico– que sustentaría diferentes clases de representaciones mentales; significa, asimismo, cosas diferentes para los partidarios de las diferentes soluciones al problema de la representación (en singular) y se plantea además en este contexto el problema de que los intentos hasta la fecha ensayados con vistas a delimitar lo representacional no alcanzan a evitar que signifique cualquier cosa, esto es, que quepa desde los mismos atribuir actividad representacional incluso a un bicicleta. Pero, más allá de esta falta de acuerdo, consideramos que la univocidad de la noción de representación que el teórico representacionalista requeriría se ve amenazada antes bien por la implausibilidad biológica que por las disputas intra e interparadigmáticas: nada invita a pensar que aquellos procesos a los que denominaríamos representacionales hayan de ser homogéneos a lo largo y ancho de árbol filogenético. Esto es, se pretende reducir la mente fenoménica a la representacional para luego

explicar ésta en términos legítimos desde un punto de vista naturalista. El problema, en pocas palabras, es que gran cantidad de fenómenos biológicos radicalmente heterogéneos podrían cumplir los principios formales planteados por los diversos teóricos representacionalistas y obedecer y plegarse, sin embargo, a mecanismos y principios explicativos enteramente diversos. En otras palabras, bien cabe que en distintos filos, y aun dentro del mismo, encontremos diferentes clases de fenómenos biológicos que el representacionalista denominaría representacionales y a los que no dudaría en atribuir sus cualidades formales favoritas pero de los cuales no quepa decir que resulten explicables en términos idénticos o análogos ni que hayan surgido y se hayan conformado de forma idéntica o análoga en el curso de la evolución. Es, en definitiva, más que probable que si «representación» significa algo en el contexto relevante –el biológico–, signifique de hecho varias cosas diferentes: el árbol filogenético tiene demasiadas ramas como para que sea de otro modo. Es decir, si «representacional» significa algo en el contexto relevante, significa del mismo modo en que lo hace «fotosensible»: apuntando a capacidades de determinados sistemas o mecanismos para producir determinadas clases de efectos que tendemos a categorizar bajo una misma etiqueta léxica. Pero se da el caso de que «fotosensible» no significa una sola cosa: existen, sin ir más lejos, células fotosensibles animales y células fotosensibles vegetales, y la explicación de cada una de ambas clases de fotosensibilidad es muy diferente –entre otras cosas porque cada una de ellas ha de atender a distintas cascadas metabólicas–. De este modo, analizar el concepto de «fotosensibilidad» suponiendo su homogeneidad e improvisar condiciones necesarias y suficientes o criterios formales de pertenencia a la clase de lo fotosensible sirve de más bien poco a la hora de allanar el camino hacia una explicación de la fotosensibilidad –entre otras cosas porque no existe *una* explicación de la fotosensibilidad–. Lo mismo, sin ningún matiz, sucede en el caso de la representación y resulta por tanto muy difícil entender de qué modo pretenden los análisis a priori que mantienen entretenidos a los teóricos de la representación contribuir a la tarea de explicar científicamente los procesos biológicos que denominan "representacionales".

Podemos calificar a nuestro segundo problema como el de la *futurología de la ciencia*. Se trata de un problema que se presentará como insalvable a cualquiera que no esté dispuesto a depositar demasiadas esperanzas en dicha "disciplina". Como apuntábamos, las especulativas teorías representacionales de la conciencia de las que nos ocupáramos en la primera parte de esta tesis necesitan dar por supuesto que la naturalización de la capacidad representacional de la mente es aproblemática, esto es, necesitan

suponer que la elaboración de *una* teoría de la representación científicamente irreprochable es un proyecto que llegará a buen puerto. Estos teóricos pretenden demostrar, pues, que la mente fenoménica puede naturalizarse, y que esto debe hacerse a través de su reducción a la mente representacional. Ellos se encargarían de trazar esquemas abstractos difícilmente operacionalizables acerca del modo de reducir aquélla a ésta y dejarían para otros la tarea decisiva, esto es, la de elaborar una teoría naturalista de la representación. La dificultad es obvia: desconocemos la ubicación de los cimientos sobre los que debiera levantarse el edificio. Tras medio siglo de discusiones nadie alcanza a ver en el horizonte la forma de una posible teoría de la representación unificada, consensuada, aproblemática y fundamental, y lo que de hecho gana pujanza en el horizonte de las ciencias cognitivas es un antirrepresentacionalismo que adopta diversas formas en su denuncia de las tradicionales nociones de computación y representación como inadecuadas y en su insistencia en la necesidad de elaborar nuevos métodos y herramientas analíticas –tomadas, por ejemplo, de la teoría de sistemas dinámicos– de cara a comprender la complejidad de las interacciones entre sistemas nerviosos, cuerpos y entornos. Cuando no sabemos de qué material está hecha la mayor de las matrioskas (la noción de representación), no podemos asegurar nada acerca del destino que correrán las que metamos dentro de ella. Obviamente, si estuviera hecha de vidrio se quebraría y dejaría de contener al resto con mayor facilidad que si estuviera hecha de acero. El problema de la futurología de la ciencia es, por otra parte, doble, por cuanto las teorías representacionales se alzan sobre dos supuestos injustificados: el primero, que algún día una teoría unitaria, fundamental y omnímoda de la representación se hallará a disposición del investigador de lo mental; el segundo, que cuando ese día llegue, esa futurible teoría fundamental habrá de sustentar, ella solita, la explicación científica de la conciencia.

El tercer problema, implícito en el anterior, consiste en que nadie ha acabado de hacer explícito el fundamento del *dogma fundamentalista* que sustenta el dogma representacionalista. ¿Cuál es el dogma fundamentalista? Un prejuicio tácito según el cual primero hay que elaborar una teoría de la representación, dado que es un fenómeno más básico, y luego elevar sobre los cimientos de la misma una teoría de la conciencia. ¿Cuál es el fundamento de este fundamentalismo? Nadie ha alcanzado a explicitarlo. La idea puede resultar tentadora, y de hecho es ella la que ha venido dando pábulo a la fatigosa labor especulativa que ha mantenido entretenidos a la mayoría de los filósofos anglosajones de la mente durante tres décadas. No obstante, insistamos, ¿cuál es el fundamento de este fundamentalismo? Según una extendida opinión, existen razones para

suponer que el heterogéneo conjunto de fenómenos biológicos que han venido denominándose representacionales son filogenéticamente anteriores a la aparición de la experiencia consciente y, por tanto, más básicos que ésta. Sin embargo, también el origen evolutivo de las proteínas y glicoproteínas que denominamos receptores celulares es anterior al de nuestros sofisticados sistemas inmunes, nerviosos y endocrinos, pero la idea de elevar el edificio explicativo de la endocrinología sobre la base de análisis apriorísticos acerca de vínculos necesarios y suficientes entre la abstracción "recepción celular" y la abstracción "endocrineidad" sólo puede ser comprendida como un mal chiste. En resumidas cuentas, el dogma fundamentalista es un dogma simplista y comprometido con la diáfana, precisa, discreta y exhaustiva referencialidad del vocabulario tradicional: asume que la economía léxica de la tradición que ha venido definiendo a lo mental como exhaustivamente descriptible mediante las etiquetas "representacional" y "fenoménico" corresponde uno a uno con la realidad de lo mental. La mente *es* representacional y fenoménica, y denominamos representacional y fenoménico a algo así como dos bolas de billar: cada una está integrada por un catálogo discreto de elementos bien delimitados dentro sus respectivos contornos esféricos, y cada una de ellas puede mantener relaciones causales con la otra y con el entorno, pero en cualquier caso, según el dogma fundamentalista, es la bola representacional la que pone en marcha a la fenoménica.

El cuarto problema ha sido ampliamente discutido en la literatura. Puede que no resulte difícil definir la experiencia consciente que tengo al mirar *La rendición de Breda* en términos representacionales, pero, ¿qué representan un súbito estado de nerviosismo inmotivado o un orgasmo? Si la solución separatista fuera correcta, el carácter cualitativo de la experiencia consciente debiera hallarse exhaustivamente constituido por contenidos representacionales, pero es ciertamente difícil imaginar un modo convincente de definir en estos términos estados mentales dotados de un indudable carácter cualitativo, como la ansiedad sin objeto (undirected anxiety), la depresión, la euforia (Searle, 1983: 2 del original, 17-18 de la traducción), el placer o el orgasmo (Rey, 1998: 441; Block, 1995: 234).

El quinto problema tiene que ver con el externismo fenoménico que, como vimos, abrazan los teóricos representacionalistas. Desde la perspectiva ortodoxa de la psicosemántica externalista, los contenidos de la experiencia visual de un pepinillo real y uno alucinado difieren –el contenido de esta última representación es, sin ir más lejos, falso o fallido– mientras, por otra parte, su carácter fenoménico puede ser exactamente el

mismo. Esta identidad en lo fenoménico acompañada de completa heterogeneidad representacional no casa nada bien con las premisas de los teóricos representacionalistas. Por otra parte, el carácter relacional que se halla a la base de este externismo obliga, como señaláramos al tratar del mismo en la primera parte de esta tesis, a conceptualizar colores, olores, sabores y también sensaciones corporales y estados de ánimo como posibles portadores de valores de verdad, lo cual vendría a implicar que el verde chartreusey (RGB 127, 255, 0) y el olor a vinagreta existen objetivamente y que un orgasmo puede ser falso. De hecho, y tal y como algunos defensores del representacionalismo han venido a admitir (Lycan, 2006a), una de las principales motivaciones para el externismo fenoménico reside en el mero hecho de que la psicosemántica que se ha impuesto para la determinación del contenido intencional es externista, de forma que si la conciencia ha de ser representacional, habrá de plegarse también a la moda en psicosemántica representacionalista.

Habiendo encontrado motivos para poner en tela de juicio que la solución separatista se muestre como una opción exenta de problemas fundamentales, pasamos a argumentar que lo mismo sucede con la solución inseparatista.

### 4._La solución inseparatista como nuevo problema

Entendemos que el modo en que los defensores del inseparatismo vinculan conciencia fenoménica e intencionalidad resulta problemático y acarrea dificultades infranqueables para una concepción naturalista de la mente. Para mostrar el sentido en que entendemos que esto es efectivamente así, comenzaremos por introducir el debate acerca de las relaciones entre intencionalidad y conciencia fenoménica criticando de forma general alguno de los elementos nucleares de las propuestas inseparatistas de cara a preparar el terreno para nuestra crítica de los conatos inseparatistas más acreditados, los pergeñados por Strawson y Searle.

Como hemos visto, la intencionalidad y el aspecto fenoménico de la experiencia consciente son actualmente concebidos como las características definitorias de lo mental, las dos caras de la moneda de lo mental. No obstante, y tal y como asimismo hemos señalado, las discrepancias en torno a la forma apropiada de establecer relaciones entre ambas características constituyen la norma en la filosofía de la mente contemporánea.

Dichas características son tratadas por unos como mutuamente implicantes y por otros como mutuamente independientes. Al seguir el hilo de las argumentaciones de los autores que tratan ambos términos del binomio de lo mental como independientes somos invitados a contemplar lo mental como constituido por dos dimensiones separables, diferenciables tanto desde el punto de vista ontológico como desde el metodológico (una representacional y otra experiencial). Cuando hacemos lo mismo con las argumentaciones que pretenden alcanzar la conclusión opuesta nos topamos con un cuadro en el que representar implica experimentar y viceversa: la intencionalidad aparece como un rasgo de los estados internos de un sistema que a) sólo puede ser adscrita al mismo en la medida en que éste sea fenoménicamente consciente, y b) el aspecto fenoménico de la experiencia consciente se nos presenta como necesariamente intencional, esto es, como portador de rasgos representacionales o, cuando menos, como en una u otra medida vinculado con tales rasgos. Cabe matizar, no obstante, que algunos de los autores que defienden la necesidad de la presencia de propiedades experienciales para la existencia de verdadera intencionalidad no estarían dispuestos a conceder que todo el reino de lo mental sea intencional o representacional, dado que, proponen, existen estados fenoménicamente conscientes que no son intencionales, como los estados de ánimo o los dolores.[2] Sea como fuere, la noción de inseparatismo ha sido recientemente elaborada por Horgan y Tienson (vid., v. g., Horgan & Tienson, 2002) para hacer referencia a este tipo de planteamiento en el que, en cualquier caso –esto es, tanto en a) como en a + b)–, la intencionalidad no aparece como una propiedad básica o autónoma. El inseparatismo es, hasta cierto punto, una perspectiva heterodoxa, por cuanto en las últimas décadas la práctica habitual en filosofía de la mente y en las ciencias cognitivas en general ha consistido en tratar a la intencionalidad y la conciencia fenoménica como ontológica y metodológicamente distinguibles e independientes. Con todo, al margen del aire fresco y la apertura de nuevos horizontes que toda heterodoxia comporta, con ellos no es suficiente: veremos a continuación, apoyándonos en las argumentaciones de Searle y Strawson, el modo en que el inseparatismo no logra hacer frente adecuadamente a una serie de problemas tan acuciantes como los que pretendía venir a resolver.

Daniel Dennett ha indicado (Dennett, 1994c: 236) que los dos temas principales de la filosofía de la mente son la conciencia fenoménica y la intencionalidad. Pocos discreparían en este punto. Dennett, al igual que el resto de los teóricos representaciona-

[2] Tal es el caso de Strawson y Searle, de cuyas propuestas nos ocuparemos en lo sucesivo.

listas, cree que la intencionalidad viene primero. En su caso, esta aseveración sería tan válida desde el punto de vista filogenético como desde el punto de vista del orden explicativo. Los inseparatistas defienden una postura radicalmente diferente: ningún estado intencional podría existir sin conciencia fenoménica. Por su parte, David Chalmers (2004) ha señalado que tanto en la filosofía moderna, de Descartes a Locke, como en los orígenes de la tradición fenomenológica (Brentano, Husserl) ambas características fueron tratadas como aspectos inseparables de lo mental. No obstante, la tendencia en la filosofía analítica de la mente ha venido siendo la contraria. Así, según la línea ortodoxa de esta tradición contemporánea, son posibles los estados intencionales no fenoménicamente conscientes. El núcleo del inseparatismo consiste en un intento de desafiar esta ortodoxia. Por nuestra parte, abordaremos a renglón seguido una concisa presentación y crítica de la tesis inseparatista cardinal: la de la necesidad de la presencia de experiencia consciente para la existencia de la intencionalidad.

Junto a la referida tesis cardinal hallamos una serie de planteamientos inseparatistas de los que no vamos a tratar aquí, entre los que cabría destacar la tesis fenomenista según la cual el contenido intencional se individualiza fenoménicamente o la propuesta fenomenológica (descriptiva) según la cual las actitudes proposicionales están dotadas de un determinado carácter fenoménico y los estados mentales típicamente conceptualizados como portadores de, exclusivamente, contenido cualitativo o fenoménico están dotados asimismo de contenido intencional. La señalada tesis cardinal de las propuestas inseparatistas sostiene, pues, que no cabe concebir estados intencionales no fenoménicos y es planteada en términos un tanto difusos: cuesta distinguir cuando se habla de una jerarquía explicativa o una filogenética. Es, por otra parte, obvio que la validez de esta tesis depende del modo en que los términos «mental», «intencional» y «fenoménico» sean definidos. Si, por ejemplo, definimos «mental» e «intencional» a la Nicolás Chauvin[3] como términos referidos a estados como los que nosotros, humanos de bien, podemos experimentar, es decir, como términos referidos a realidades que cobran sentido sólo a la luz de la fenomenalidad de la conciencia humana, en la propia definición de los términos se hallan ya las conclusiones a las que pretendíamos arribar –se trataría de una táctica argumentativa circular que podemos denominar *conclusiones-en-las-definiciones, luego premisas, luego, nuevamente, conclusiones*–. En dos palabras: si, como no resulta inusual, la intencionalidad es calificada de *verdadera* u *original* sólo

[3] Para una definición amplia de «chauvinismo» en este contexto vid. Shani (2008).

cuando se trata de intencionalidad fenoménica humana, es decir, sólo cuando se trata de estados mentales que refieren a sus objetos tal y como yo siento que los míos refieren, entonces, obviamente, el uso del término queda restringido desde el principio, siendo reservado desde ese mismo momento para su aplicación a los estados mentales cuya naturaleza estamos tratando de desentrañar.

Las anteriores alusiones al chauvinismo implícito en la definición de intencionalidad propia de posturas inseparatistas requieren de una elaboración que permita superar el aire de acusación trivial que las mismas han aportado a la crítica del inseparatismo que aquí emprendemos. No se trata simplemente de que la concepción de la intencionalidad que ofrece el inseparatista resulte insatisfactoria por cuanto niega la posibilidad de adscribir intencionalidad a organismos habitualmente considerados incapaces de sostener estados mentales conscientes como estos que *yo* experimento e intuyo que *tú* experimentas, sino –también y más bien– por cuanto el núcleo del inseparatismo consiste en este punto en definir la característica básica de cualquier sistema que pueda considerarse representacional en términos de propiedades que no hay motivos ni teóricos ni empíricos que impidan considerar como filogenéticamente posteriores –esto quiere decir que nada niega a priori la posibilidad de que dentro del reino animalia los estados internos de algún miembro de la familia arthropoda fueran referenciales antes de que ningún vertebrado alcanzara a ser fenoménicamente consciente–. Una interpretación directa de este aspecto del inseparatismo, una interpretación que atienda al modo en que dicha postura describe *propiedades representacionales naturales* en general –tomamos esta expresión de Geisz (2009), pero aludimos con ella a propuestas del tipo de la de Millikan (vid., v. g., Millikan, 2001)– antes que al modo en que es aplicable a una descripción fenomenológica de la mente consciente humana en particular, una interpretación tal, pues, obliga, desde la perspectiva filogenética, a concebir la irrupción de experiencia consciente como condición necesaria para la adscripción de intencionalidad a cualesquiera estados de cualesquiera organismos. Esto, a su vez, fuerza a la elección entre 1) alguna suerte de panfenomenalismo ("sí –diría el suscriptor de esta opción–, esa referencialidad pre-humana es *verdaderamente intencional*, dado que la experiencia consciente es un rasgo fundamental de la naturaleza que hallamos en todos y cada uno de sus órdenes y niveles"),[4] 2) una definición estrecha de intencionalidad (calificada entonces

[4] Haciendo caso omiso de algunos de los más tenaces resortes de una recta sindéresis, y creando, además, el nuevo problema de explicar los mecanismos por los cuales la conciencia ordinaria surge de las protoconciencias –*ex hypothesi*– diseminadas por todo lo ancho y largo del mundo natural (Lycan, 2006b),

de "original" o "verdadera") según la cual la referencialidad que acaso da la sensación de que encontramos en organismos o sistemas que tenderíamos a clasificar como inconscientes no tiene nada que ver con la "verdadera intencionalidad" (la humana *virtus representativa*), o 3) una concepción de lo mental según la cual todas las propiedades mentales surgieron juntas de golpe y porrazo: nada fue una representación de nada hasta el momento en que alguien sintió o experimentó que ése era el caso, esto es, nada fue mental hasta que surgió esto que entendemos por mente, esto que experimentamos como mente, y entre sus plausibles precursores, nada mereció tal nombre porque tan siquiera se trataba de estados de los que cupiera afirmar que fueran sobre algo.[5] "Lo mental –se dice el inseparatista– es *esto*: con una mente como esta mía surge la mente". Al igual que en Descartes, la auto-ostensión es la más acabada entre las definiciones de «mente» que ofrece el inseparatista, del mismo modo que una súbita explosión cuvieriana, con antecedentes esencialmente heterogéneos, parece ser la mejor de las versiones de la génesis de lo mental que puede ofrecernos el Plinio (*Naturalis historiae*) inseparatista –siempre que no se sienta tentado por la metafísica panpsiquista.

Asistidos por la orientación que las indicaciones tímidamente bosquejadas con lo antedicho puedan ofrecernos, centrémonos ya en la crítica de las propuestas inseparatistas de Strawson y Searle.

### 4.1._El inseparatismo de Strawson

La primera acotación que hemos de realizar antes de iniciar nuestra crítica de las tesis de Strawson es que las mismas no nos son presentadas –explícitamente– como inseparatistas. Sin embargo, Strawson se compromete con la tesis cardinal del inseparatismo, el segmento crucial de esta concepción del vínculo entre intencionalidad y conciencia fenoménica, un segmento que proponemos denominar *el mito de la verdadera intencionalidad*, es decir, no ya la idea de que la mente humana, y sólo ella, posee *in-*

---

Strawson (vid. Strawson, 2006) ha venido a desembocar en esta opción, que a su vez le ha conducido a la defensa de un cierto pre-epifenomenalismo: "there's no good reason (…) to think that [consciousness] first came on the scene because it had survival value. Natural selection needs something to work on and can only work on what it finds. Experience/consciousness had to exist before it could be exploited and shaped, just as non-experiential matter did. The task of giving an evolutionary explanation of the existence of consciousness is exactly like the task of giving an evolutionary explanation of the existence of matter: there is no such task. Natural selection moulds the phenomena of experience it finds in nature" (Strawson, 2010: 304).

[5] Téngase presente la habitual identificación de «intentionality» con «aboutness», crudamente presentada por Dennett & Haugeland (1987) y asumida por Strawson –asunción "matizada" en el apéndice de la segunda edición de *Mental Reality* (Strawson, 1994/2010).

*tencionalidad original* –una idea que puede leerse en Searle (1992; 2004a)–, sino la de que la conciencia fenoménica constituye una condición de posibilidad de la intencionalidad en general.

Decíamos que Strawson no se presenta como inseparatista, y es que la suya no es una propuesta inseparatista completa: a pesar de haber defendido la existencia de *cognitive/understanding-experiences* (Strawson, 1994/2010: 4 y ss.) y por tanto la fenomenalidad de tales estados intencionales, Strawson entiende que no toda experiencia dotada de aspecto fenoménico debe ser catalogada como intencional. Así, apunta, muchos estados mentales no parecen tener un objeto definido o dirigirse a él (sensaciones corporales y estados anímicos son los ejemplos habituales en este punto) y nadie dudaría de su carácter mental. Independientemente de la pertinencia de esta propuesta en tanto que planteamiento fenomenológico (descriptivo), Strawson extrae de la misma una conclusión que no argumenta (pero que hemos de suponer implícitamente sustentada por el conjunto de su argumentación): la intencionalidad no es la clave para desentrañar lo mental, dado que, según el británico, resulta desorientador presumir que solucionar el *supuesto problema* que comporta ofrecerá la clave para una adecuada comprensión de la mente (Ibíd.: 177).

Strawson asume que los postulados naturalistas y fisicalistas –nociones que usa indistintamente (vid., v. g., Strawson, 2005)– en filosofía de la mente son correctos, y es precisamente partiendo de esta asunción que presenta el problema de la intencionalidad como un *supuesto* problema a través de la que denomina *tesis de la ausencia de problema*. Según la misma, de cara a ofrecer una explicación naturalista de la mente, la intencionalidad no se muestra problemática, es decir, que no hay en ella un problema profundo o desconcertante *distinto del problema de la experiencia*. En otras palabras, la intuición[6] que Strawson se propone articular pivota en torno al siguiente eje: ofrecer una explicación de la intencionalidad no supone una amenaza para una concepción fisicalista de le mente, al menos no una distinta de la que supone la experiencia consciente.[7]

[6] Hablamos de *intuición* dado que, ya en la primera edición de *Mental Reality*, Strawson plantea esta idea en esos términos, y lo hace en un tono humilde que se verá matizado, por ejemplo, en el apéndice a la segunda edición –y no sólo en él, sino también en la propia primera edición, conforme avanza su argumentación en el capítulo séptimo–. Así, encontramos en la referida primera edición una declaración de intenciones muy poco ambiciosa: Strawson nos dice que se limitará a examinar una *intuición* acerca de cuya pertinencia, asegura, no pretende convencer a nadie (Strawson 1994/2010: 178).

[7] La tesis, así formulada, consta de dos partes, una primera bastante inocua en apariencia (la intencionalidad no supone un problema para el fisicalismo), y una segunda que viene a sostener exactamente lo contrario: la experiencia consciente desafía el marco fisicalista o se muestra dentro del mismo problemática y, tal y como esta segunda parte de la tesis sugiere, como la experiencia consciente tiene algo que ver con la intencionalidad, entonces ésta, ahora sí, se muestra problemática dentro de un marco naturalista, pero

Strawson se muestra moderado en primera instancia: expone en las primeras páginas de su primera aproximación al vínculo entre intencionalidad y experiencia consciente su modesta intención de, meramente, examinar una intuición. Declara asimismo en ellas no tener litigio abierto con quienes defienden que la experiencia consciente no es necesaria para la intencionalidad o no es una condición de la misma y, por último, afirma que su principal interés será el de exponer la tesis de la ausencia de problema, el de desarrollar una mera intuición. Intenciones nada ambiciosas, o al menos ésa es la primera impresión porque, conforme avanza su argumentación y vamos encontrándonos con expresiones adjetivas de semántica nada inocente –como intencionalidad *original*, *genuina*, *auténtica*, *verdadera* o *intrínseca*–, comenzamos a poner en cuestión la veracidad de tales declaraciones de intenciones, cosa que dejamos de hacer –para pasar a dudar decididamente de ella– al leer las últimas páginas del séptimo capítulo de *Mental Reality*, momento en que asistimos a un viraje que lleva al moderado autor que escasas páginas atrás hacía gala de una circunspecta prudencia al admitir la imposibilidad de probar la necesidad de la experiencia consciente para la intencionalidad a afirmar que la experiencia es una *condición necesaria* de la auténtica referencialidad (*genuine aboutness*) (Strawson, 1994/2010: 211) o que no se puede tener intencionalidad si no se es un ser experienciante (Ibíd.: 208).

Strawson lleva a cabo el salto argumentativo al que acabamos de aludir atravesando una enmarañada red de distinciones, análisis conceptuales y experimentos mentales que, incluso desde la más caritativa de las interpretaciones, no ofrece soporte ni aporta legitimidad al mismo. Así, nos encontramos con una injustificable dualidad argumentativa: la completa heterogeneidad, la flagrante disrupción ilativa habida entre las consecuencias extraídas[8] y el lugar del que pretenden extraerse.[9]

---

no por sí misma, es decir, que no se muestra problemática en tanto que intencionalidad, sino por cuanto ella es, de algún modo, dependiente de la experiencia consciente.

[8] Es decir, la tesis de la ausencia de problema y la necesidad de la experiencia consciente para la existencia de la intencionalidad, necesidad que –después de todo, es decir, a pesar de la insistencia de Strawson en que su interés central en el texto al que estamos haciendo referencia reside en la articulación y defensa de la tesis de la ausencia de problema– se revela como objetivo inconfesado del autor.

[9] Es decir, la referida red de análisis, distinciones –entre tipos de intencionalidad, como, por ejemplo, intencionalidad hacia objetos existentes concretos o intencionalidad hacia objetos abstractos no existentes– y experimentos mentales, una red de la que, incidamos en ello, resulta imposible extraer las consecuencias que Strawson parece pretender. Todo lo más que cabe defender es que la misma ofrece intuiciones acerca de extremos discutidos en la filosofía analítica de la mente en relación con el *contenido estrecho* de los estados mentales al hablar de *adultos neonatos o instantáneos* que resultan ser *cerebros en cubetas* y engendros filosóficos similares, los cuales, como indicábamos, ciertamente iluminan aspectos interesantes de la discusión acerca del llamado *contenido estrecho*, pero, se mire por donde se mire, poco o nada dicen acerca del vínculo intencionalidad-fenomenalidad que Strawson intenta trazar.

La señalada maraña de análisis y experimentos mentales no juega un papel muy definido dentro de la argumentación de Strawson, al menos ninguno más allá del del pretexto para desarrollar sus intuiciones en el contexto de diversos artefactos narrativos. No comporta, pues, la referida maraña soporte argumental alguno, sino sólo una iteración: Strawson tiene una intuición que inserta de diversos modos en diferentes relatos. No obstante, consideramos que resultará justo y esclarecedor prestar atención a algún ejemplo concreto para comprender la orientación de la argumentación de Strawson en su aproximación a la cuestión del vínculo intencionalidad-experiencia. Así, por ejemplo, Strawson nos propone que consideremos un ser sin experiencia: un misil termodirigido que sigue su blanco. ¿Posee intencionalidad semejante dispositivo? Algunos dirían cosas como que su intencionalidad es *derivada* de la de sus diseñadores, pero Strawson plantea la cuestión desde el punto de vista de los "artefactos no-artificiales": consideremos que el misil surge ya programado por un extraño azar cósmico y hagámonos la misma pregunta. Si atribuimos referencialidad a la representación de su objetivo que tendría este "artefacto no-artificial", entonces separamos tajantemente la cuestión de la intencionalidad de la de la experiencia –por cuanto le adscribimos aquélla aun cuando le negamos ésta–, con lo cual ofrecemos respaldo a la tesis de la ausencia de problema. Pero no parece que a Strawson le haga especial ilusión dicho respaldo cuando enseguida se deshace del mismo para sugerirnos que si estamos dispuestos a atribuir intencionalidad a semejante dispositivo acabaremos atribuyendo intencionalidad a cualquier cosa, incluso a un estanque que refleje cualquier imagen en su superficie, y al deshacerse de este modo del referido respaldo pone de manifiesto que la tesis que verdaderamente le interesa defender no es la de la ausencia de problema, sino la de la necesidad de la experiencia consciente para la intencionalidad. Se trata de una estrategia verdaderamente extraña: *dice* una cosa pero *muestra* otra bien diferente –y su mostrarla culmina extravagantemente con la inclusión en la segunda edición de *Mental Reality* de "Real intentionality 3", un texto (publicado anteriormente en Strawson, 2008), que hace explícitamente evidente que la idea que a Strawson le interesa defender es la que hemos denominado tesis inseparatista cardinal.

En definitiva, Strawson utiliza estos experimentos mentales basados en "artefactos no-artificiales" (como robots surgidos espontáneamente por caprichosos albures cósmicos u ordenadores irradiados) para hacer plausible su intuición de que no podríamos atribuir intencionalidad a tales "artefactos" más que metafóricamente, pues sólo reaccionan de determinadas maneras que podemos interpretar como signos de intencionali-

dad, una intuición que descansa sobre la idea de que la diferencia entre su irreferencialidad y nuestra referencialidad estriba sólo en la experiencia, la cual, propone Strawson, *sin duda marca la diferencia* (Strawson 1994/2010: 191). El modo en que discurre su argumentación haciendo uso de tales experimentos mentales habrá dejado perplejos a no pocos lectores, pues en la misma las premisas no difieren de las conclusiones. Strawson argumenta, exactamente, que la diferencia entre la irreferencialidad de sus "artefactos no-artificiales" y nuestra referencialidad estriba sólo en la experiencia, la cual, *sin duda*, marca la diferencia. Es decir, que la experiencia marca la diferencia porque sin duda marca la diferencia y el espacio-tiempo es curvo, exactamente, porque sin duda es curvo. Conclusiones en las definiciones, luego premisas, luego, nuevamente, conclusiones. Además, «sólo reaccionan de determinadas maneras» se traduce en el léxico strawsoniano por *intencionalidad conductual*, de la cual todo lo que nos dice el filósofo británico es que no es *verdadera* intencionalidad, pues, independientemente de la refinación de nuestra definición de conducta (la cual, nótese, podría incluir desde el más nimio de los detalles relacionados con la fotoquímica de las opsinas al más completo y marcadamente dinámico diagrama de interacción entre los geniculados laterales, las diferentes áreas de la corteza occipital con aquéllos y entre sí, etc.), faltará siempre el ingrediente esencial: la experiencia, sin la cual cualquier sistema podrá, meramente, comportarse *como si* tuviera intencionalidad. Nuevamente las conclusiones son, desconcertantemente, idénticas a las definiciones de los términos en que son enunciadas las premisas.

Por otra parte, aprovecha Strawson estos ejemplos para insistir en que lo que le importa es su tesis de la ausencia de problema, para mostrar acto seguido que lo que le interesa realmente es el modo en que la experiencia es necesaria para la intencionalidad. Así, afirma que independientemente de que atribuyamos o no intencionalidad a tales "artefactos no-artificiales", la tesis queda a salvo: si lo hacemos, la intencionalidad no es misteriosa –pues puede entenderse entonces que ella es sólo cuestión de ensamblaje mecanicista–, y si no lo hacemos, entonces concedemos que la intencionalidad es misteriosa sólo por su inextricable relación con el problema de la experiencia, y esta segunda opción, por algún motivo, parece ser la que más atrae a Strawson. A esto añade Strawson 1) que muchos se sienten inclinados a decir que mientras que sus pensamientos se refieren verdaderamente a algo, los de una "máquina" (en léxico strawsoniano, un *ser no experiencial*) no, 2) que muchos se sienten inclinados a decir que sólo en el caso experiencial se da verdadera intencionalidad, 3) que ésta es una intuición de gran alcance, y 4) que lo único que la experiencia añadiría al caso de los seres no experienciales

sería la propia experiencia. De este modo, concluye –una conclusión que, nueva y especialmente, le viene grande a sus premisas–, la intencionalidad consiste en experiencia sumada a todas las propiedades, capacidades y disposiciones no experienciales que nosotros (locus de la *verdadera* intencionalidad) compartimos con las máquinas o seres no experienciales. En este sentido, Strawson ha llegado a asegurar que "las únicas entidades verdaderamente intencionales son los episodios experiencialmente conscientes" (Strawson, 2005: 41τ), de donde concluye que la intencionalidad es un fenómeno esencialmente experiencial y que los estados disposicionales (como creencias no ocurrentes) ni son intencionales ni pueden tener contenido intencional, lo que le lleva de vuelta a su intuición de que no pueden hallarse verdaderos fenómenos intencionales donde no hay experiencia (Ibíd.: 60), intuición en la que se basa para sentenciar que la ocurrencia de un estado mental y, a fortiori, la ocurrencia de un estado intencional puede acaecer, exclusivamente, en un sujeto capaz de sostener experiencias conscientes, cosa que, según su opinión, no requiere de ulteriores justificaciones dado que es un hecho obvio acerca de lo que la palabra «mental» significa (Ibíd.: 46).

La demostración strawsoniana de la tesis inseparatista cardinal tendría pues la siguiente forma: la experiencia consciente es una condición necesaria de la intencionalidad porque sin duda lo es. Dejando en manos del lector la valoración del alcance apodíctico y la refinación erística de una tal demostración pasamos ya a ocuparnos de la ensayada por Searle.

### 4.2._El inseparatismo de Searle

El inseparatismo de Searle gravita sobre su *Principio de Conexión*, que postula: "la adscripción de un fenómeno intencional inconsciente a un sistema implica que el fenómeno es en principio accesible a la conciencia" (Searle, 1990b: 586τ). El principio de conexión, tal y como Juan Hermoso y Pedro Chacón han acertado a señalar, "ocupa un lugar central en la teoría de la mente de Searle" (Hermoso & Chacón, 2000: 172). No en vano, dicho principio constituyó el eje en torno al cual girara la reformulación searleana de la hipótesis del Trasfondo, una reformulación muy significativa, dado que con ella, como veremos en la tercera parte, la concepción searleana del inconsciente encuentra una formulación definitiva en la cual la noción de mente o intencionalidad inconsciente –y con ella toda noción de lo mental– se encuentra supeditada a la de mente consciente.

El principio de conexión fue presentado por vez primera en Searle (1989a), y después en Searle (1990b) y Searle (1992). En esta última versión, el principio es anunciado como la idea de que todo estado intencional inconsciente es o ha de ser para contar como tal *en principio* accesible a la conciencia. Searle prepara en la misma el terreno para la aquiescencia mediante una serie de artefactos retóricos que hacen las veces de exordio a su "argumento" –pues el principio de marras tiene la forma de un argumento aunque, pide Searle, no debe ser comprendido como una deducción a partir de axiomas–. Destacamos aquí entre los artefactos aludidos el hiperbólico "nado valientemente a contracorriente": Searle acusa –subrepticiamente, cabría decir, pues ni cita ni comenta explícitamente ninguna propuesta concreta– a todo intento –separatista– de naturalización de la intencionalidad de formar parte de una taimada ortodoxia cuya finalidad es la de construir una ciencia objetiva de la mente a la que le sea dable rehusar enteramente toda apelación a la experiencia consciente, y frente a tal ortodoxia aparece él en su relato, alzando la voz en solitario. El siempre elocuente Daniel Dennett describe esta autopercepción de Searle en los siguientes términos: "[Searle] sees himself as an iconoclast, waging lonely battle against 'the tradition' –the 'mainstream orthodoxy' of functionalistic materialism that has unjustly captured the flag of the scientific establishment" (Dennett, 1993a: 193).[10]

Searle (1992: 153 del original, 161 de la traducción) constata que en décadas recientes se ha realizado (desde de las filas de la señalada ladina ortodoxia) un esfuerzo considerable por llevar a cabo un proyecto que considera desorientado y ofuscador: el de separar intencionalidad y conciencia e intentar dar cuerpo a una teoría de la intencionalidad en la cual la experiencia consciente ni entre ni salga. "La idea es tratar la intencionalidad 'objetivamente', tratarla como si los rasgos subjetivos de la conciencia no importaran realmente" (Ibíd.). Un tratamiento análogo de la intencionalidad da cabida a estados y procesos mentales que de ningún modo se asemejan a los estados y procesos mentales que nosotros, humanos, experimentamos. Y a Searle eso no acaba de parecerle del todo razonable, porque, según su *voto* –acepciones primera, tercera y, particularmente, octava de dicha entrada en la vigésima segunda edición del diccionario de la lengua española de la RAE–, y dramatizando sus consecuencias últimas, tener mente, ser sujeto pasible de adscripciones y predicados mentales, es tener *esto* que los humanos

[10] La tendencia a hacer frente a la "ortodoxia filosófica", como ha señalado Luis M. Valdés (Valdés, 2009: 578), ha sido una constante en la carrera de Searle. No resulta pues descabellada la suposición de que este autoconcepto comenzara a fraguarse ya cuando, recién licenciado, publicara un artículo (Searle, 1964) desafiando la falacia naturalista mooreana.

*experimentamos*. De este modo, Searle desacredita la posibilidad de intencionalidades y mentalidades subpersonales rebatiendo entre líneas (ni cita, ni nombra, ni argumenta explícitamente al respecto) propuestas en psicología de la percepción del tipo de la inferencia inconsciente (que se remonta a mediados del siglo XIX, en concreto, al *Handbuch der Physiologischen Optik* que Herman von Helmholtz publicara entre 1856 y 1867) o del tipo de las reglas sintácticas innatas de la psicolingüística de inspiración chomskiana. Searle no cita ningún trabajo concreto, pero alude a este tipo de planteamientos en psicología de la percepción y en psicolingüística por lo que ellos comparten, a saber, el recurso a estados y procesos representacionales que pretenden ser mentales pero que nada tienen que ver con lo que experimentamos como mental sino que, de hecho, tienen vetado el acceso a la conciencia, es decir, por su adscripción implícita a una tesis según la cual la intencionalidad es ontológica y metodológicamente independiente de la experiencia consciente. Asimismo, en Searle (1990b: 589) encontramos que, nuevamente sin hacer referencia o rebatir de forma específica ninguna propuesta teórica concreta, el de Denver desacredita la posibilidad de la existencia de *modelos mentales* –en una alusión obviamente dirigida antes a Jonhson-Laird (1983a) que a Craik (1943)–, de *imágenes 2D* –en clara alusión al pionero modelo computacional de la visión de David Marr– o de un *lenguaje del pensamiento* –en clara alusión a Fodor (1975).

Hechos los votos, puestas las cartas boca arriba, Searle pregunta: ¿qué es lo que hace que algo sea mental aun cuando no es consciente? Por toda respuesta: que de algún modo podría ser consciente. Esta respuesta *es* el principio de conexión. Antes de entrar en materia, Searle dice que tenemos "profundas" razones para creer que, efectivamente, las cosas son como esta respuesta las pinta. Pero hay también un argumento a su favor, a favor de esta escueta respuesta a aquella enorme pregunta, a favor de esa respuesta que Searle llama "Principio de Conexión". Nos atendremos en nuestra exposición del mismo a la versión comprimida –en siete puntos, frente a los diez de la versión de 1989– y definitiva contenida en Searle (1992). Al igual que en la versión de 1990, y tal y como ya entonces señalara, el argumento es idéntico y, a pesar de los cambios, "su estructura básica se ha mantenido intacta" (Searle, 1990b: 596τ) desde la primera versión hasta la definitiva, en la que, como apuntábamos, nos centraremos.

Uno de los motivos –una de esas profundas razones– que Searle alega como pretexto para la formulación de su principio de conexión es que no podemos habérnoslas sin el poder explicativo del inconsciente (Searle, 1992: 151 del original, 159 de la traducción). No podemos prescindir de él, pero podemos –y es ésta, precisamente, la in-

tención de Searle– plantear la noción del inconsciente como parasitaria de la noción de conciencia. En este sentido, Searle "intentará replantear la relación entre la conciencia y el inconsciente para mostrar (…) la centralidad de la conciencia con relación a lo inconsciente y la dependencia de la noción de inconsciente de la noción de conciencia" (García Valero, 2003: 67). Searle considera además perentorio que en la disquisición de la intersección entre intencionalidad e inconsciente nuestra concepción del inconsciente nos permita 1) respetar la distinción entre estados genuina o intrínsicamente intencionales –sintomáticamente, Searle usa alternativamente la expresión anterior y esta otra: «intrínsicamente mentales»– y estados que no lo son sino derivadamente, y 2) conservar también para los estados mentales inconscientes la aspectualidad inherente a toda intencionalidad: cada estado intencional representa sus condiciones de satisfacción solamente bajo ciertos aspectos, los cuales tienen que importar al agente –Searle resume este segundo punto diciendo que todo estado intencional tiene un cierto *contorno de aspecto*, cosa que, congruentemente, deberán tener asimismo los estados mentales inconscientes–. Son estos dos extremos, según el de Denver, los que proporcionan las bases para el argumento que ha de soportar la idea que Searle denomina "principio de conexión".

Recogemos a continuación los siete puntos de los que consta el señalado argumento (Searle, 1992: 156-160 del original, 164-169 de la traducción), no sin antes incidir en el significado del principio de conexión al que dicho argumento sirve, significado que, con Itay Shani, podemos resumir como sigue: "every unconscious intentional state is *potentially conscious*" (Shani, 2007: 66):[11]

1._ *Hay una distinción entre intencionalidad intrínseca e intencionalidad como-si; sólo la intencionalidad intrínseca es genuinamente mental*.

2._ *Los estados intencionales inconscientes son intrínsecos*.

3._ *Los estados intencionales intrínsecos, ya sean conscientes o inconscientes, tienen siempre contornos de aspecto*.

4._ *El rasgo del aspecto no puede caracterizarse sólo, de manera exhaustiva o completa, en términos de predicados de tercera persona, conductistas, o incluso neurofisiológicos*. Searle afirma que la evidencia de tercera persona deja indeterminado el carácter de aspecto de los estados intencionales y asegura que habrá siempre entre la ontología del aspecto y los fundamentos epistémicos objetivos un vacío inferencial. Se-

[11] Énfasis en el original.

gún Searle, pues, el contorno de aspecto de un estado intencional no puede ser descrito adecuadamente en términos de predicados de tercera persona, conductuales o neurofisiológicos (Searle 1990b: 587). De ahí que, desde su punto de vista, ninguna acumulación de datos neurofisiológicos, objetivos, por cuantiosa que pudiera llegar a ser, ofrecerá "datos de aspecto". Así, hemos de suponer, tan siquiera una neurobiología hipotéticamente completa y correcta permitiría determinar si mi pensamiento consciente sobre Héspero se refiere al lucero vespertino o al matutino, o si mi deseo de agua lo es de $H_2O$, cosa que requeriría de una inferencia de lo neurofisiológico a lo intencional que haría explícito el modo en que la especificación de lo neurofisiológico en términos neurofisiológicos no es aún una especificación de lo intencional.

5._ *Pero la ontología de los estados mentales inconscientes, en el momento en que son inconscientes, consiste enteramente en la existencia de fenómenos puramente neurofisiológicos*. En este punto surge una contradicción, pues si la ontología de la intencionalidad inconsciente consiste enteramente en hechos neurofisiológicos, objetivos, de tercera persona, y, al tiempo, dicha intencionalidad posee el referido rasgo de aspecto, que no puede estar constituido por tales hechos, necesitamos (Searle necesita) añadir un ingrediente a esa ontología, y Searle, en un giro que trae a las mientes el escolástico –hemos de remontarlo en verdad a *De Anima* (II, 1, 412 a 19)– "corporis physici organici potentia vitam habentis" (acendrado de perifollos: "cuerpo con vida en potencia"), repentiza el siguiente:

6._ *La noción de un estado intencional inconsciente es la noción de un estado que es un posible pensamiento o experiencia consciente*. Según Searle, para que los fenómenos mentales inconscientes sean genuinamente mentales, y, en sus términos, intrínsecamente intencionales, tienen que conservar de algún modo contornos de aspecto, y el único modo que se le ocurre a Searle es postulando que esos fenómenos inconscientes son de hecho posibles estados fenoménicamente conscientes, que la intencionalidad de los mismos depende, pues, de su potencial fenomenalidad. Searle acusa que la noción problemática es en estas frases la de posibilidad (él la encuentra problemática porque un estado intencional del tipo de los que anda buscando, entiende, podría tener, después de todo, impedido –*imposibilitado*– su acceso a la conciencia)[12], pero todo lo que hace

[12] Searle habla de las causas de un impedimento semejante en términos de lesiones cerebrales o represiones psicológicas, a pesar de lo peregrino de la primera idea (que lleva a pensar en un ingenuo modelo espacial de la conciencia como el que Searle critica; así, cuando impugna el modelo de los peces-estados-mentales que nadan en el inconsciente y emergen idénticos a la superficie parece tener en mente algo parecido a lo que sugiere esta idea de un estado neurofisiológico-mental que cuenta con todo lo que hay

respecto de dicho carácter problemático es remitir al último punto de su argumento (el séptimo) y presentarlo como una explicación adicional de éste (el sexto) implicada, además, por su conjunción con el anterior.

7._ *La ontología del inconsciente consta de rasgos objetivos del cerebro capaces de causar pensamientos conscientes subjetivos*. Al describir algo como un estado intencional inconsciente, dice Searle, caracterizamos una ontología objetiva en virtud de su capacidad causal de producir conciencia fenoménica. Esta oscura frase, que transcribimos casi literalmente, parece dirigida a resolver la dificultad que implicaría una virtual imposibilidad para acceder a la conciencia por parte de un estado potencialmente consciente: con ella se nos presenta la causalidad aquí operante en términos disposicionales. Tales disposiciones seguirían existiendo a pesar de un "bloqueo" del tipo de los que Searle contempla (vid. supra: última nota al pie) al igual que un volumen dado de cloruro sódico seguiría siendo soluble aun cuando jamás se humedeciera. "El concepto de intencionalidad inconsciente es entonces el de *latencia* relativa a su *manifestación* en la conciencia"[13] (Searle, 1992: 161 del original, 169 de la traducción), concluye Searle.

Searle resume su argumento y los motivos que le condujeran a la formulación del mismo como sigue. La noción de inconsciente (concebida por él como el conjunto completo de las actitudes proposicionales que un sujeto tiene mientras no es consciente de ellas) es valiosa desde el punto de vista explicativo. Hemos pues de preservarla. Pero, mientras un estado mental permanece inconsciente (incidimos en que su modelo es aquí el de las actitudes proposicionales) no hay nada respecto del mismo a parte de neurofisiología. ¿Cómo entenderlo entonces como intencional –y si, como en su modelo, nos hallamos ante actitudes proposicionales han de serlo forzosamente– cuando el contorno de aspecto es por un lado inherente a todo estado intencional y por otro permanece ausente en el nivel neurofisiológico? La solución de Searle consiste en postular que el contorno de aspecto puede ser atribuido a las estructuras neurofisiológicas pertinentes sólo si entendemos que éstas tienen la capacidad de producir estados conscientes. La diferencia entre los estados neurofisiológicos que nada tienen de mentales y los estados neurofisiológicos inconscientes pero mentales estaría de este modo en que estos últimos

---

que tener para acceder al reino de la mentalidad consciente, excepto porque se topa con que han cerrado el camino en algún punto del trayecto neuronal hacia la superficie) y del carácter fuertemente psicoanalítico de la segunda –no hay que perder de vista que Searle contrapone su concepción del inconsciente tanto a la propia de la ortodoxia cognitivista como a la propia de la tradición psicoanalítica.

[13] Las cursivas son de Searle.

son candidatos para la conciencia. Lo mental se mostraría así como neurofisiología más conciencia fenoménica (bien se halle ésta en acto o en potencia), a pesar del modo en que Searle concibe como idénticas ambas nociones (Searle, 2004a: 124). La conclusión es, en cualquier caso, que los estados mentales *profundamente inconscientes* o por principio inaccesibles a la conciencia, tal y como avanzábamos, no existen, siendo así que, en la propuesta de Searle, "la noción de estado mental inconsciente implica necesariamente accesibilidad a la conciencia, y carece por tanto de sentido la postulación llevada a cabo por científicos cognitivos y filósofos de la mente de estados mentales que, en principio, son inaccesibles a la conciencia *y no guardan relación con la experiencia subjetiva*" (Hermoso & Chacón, 2000: 175).[14]

El principio de conexión ha sido criticado desde diversas perspectivas que, entendemos, dados los fines de nuestra argumentación –y, por añadidura, por mor de la concisión– no resulta ni pertinente ni particularmente interesante traer en este ínterin a colación. Nos interesa aquí, nada más, hacer notorias las dificultades a las que no puede hacer frente el inseparatismo searleano en su intento de probar la necesidad de conciencia fenoménica para la existencia de intencionalidad. A tal fin, comenzaremos por esclarecer brevemente el enredo conceptual que sustenta el argumento de Searle. Searle pretende convencernos de que la intencional depende de lo fenoménico, pero en su argumento (pasos 3 y 4) se introduce la noción de contorno de aspecto como una especie de noción de comodín. "El argumento de Searle depende de la afirmación de que sin conciencia, no hay manera de explicar el contorno de aspecto que exhibe la intencionalidad" (Chalmers, 1996: 334 del original, 360 de la traducción). Como el contorno de aspecto es, según Searle, esencial para la intencionalidad, y como el mismo no puede estar ausente en ningún estado fenoménicamente consciente, la conclusión de Searle está explícitamente contenida en las definiciones de las nociones que involucra en sus premisas. En este sentido, ha llegado a afirmarse que la definición habitual de "conciencia fenoménica" no parece distar mucho de la que Searle ofrece de la noción de "contorno de aspecto": "the definition of phenomenal consciousness looks to be a close cousin of Searle's notion of 'aspectual shape'" (Gunson, 1998: 156). En la argumentación de Searle, sólo lo fenoménico podría proporcionar a la intencionalidad su característico contorno de aspecto, pero, al constituir éste una característica tanto de la inten-

[14] Las cursivas son nuestras.

cionalidad como de la conciencia fenoménica, el argumento podría funcionar perfectamente en ambas direcciones. Si aceptamos la pertinencia de los términos en que Searle plantea la cuestión, aceptando así discutirla desde ellos, nada más allá de asunciones previas no tematizadas puede favorecer uno u otro orden explicativo o constitutivo.

En un marco más general, Searle defiende la idea de que "sólo un ser capaz de conciencia puede tener estados intencionales" (Searle, 2006: 104τ). Pero en este marco más amplio el problema persiste. En pocas palabras, nadie parece disponer del modo de elucidar en qué medida los fenómenos mentales son intrínsecamente intencionales a causa de lo fenoménico o viceversa. El orden de la explicación puede funcionar, si es que funciona, en ambos sentidos: la presencia de características intencionales puede explicar la presencia de características fenoménicas y viceversa (Gertler, 2001).[15] En definitiva, y como en otro contexto dijera Feyerabend, todo vale. Podemos preguntarnos, entonces, si cabe probar que las características intencionales o las fenoménicas son, unas u otras, previas o básicas desde un punto de vista explicativo o constitutivo, pero parece que no nos llevará muy lejos nuestro inquirir, ya que, por lo pronto, nadie ha alcanzado a ofrecer nada parecido a una demostración en uno u otro sentido mientras todos los conatos emprendidos en esta dirección han consistido en explicitaciones de mayor o menor profundidad de las intuiciones y asunciones previas del teórico empeñado en convencernos de la justedad de una u otra dirección explicativa y nunca en nada parecido a sugerencias para, verbigracia, interpretar u obtener datos empíricos. Por lo que, justamente, a la interpretación de la evidencia empírica respecta, cabe notar que la visión ciega (una panorámica resoluta y actual de manos del descubridor del fenómeno puede leerse en Weiskrantz, 2007)[16] o el más simple paradigma experimental de *priming*[17] ofrecen datos difícilmente articulables dentro del marco searleano, dado que ha-

[15] Cabe aplicar esta idea a argumentos análogos, como ése que Siewert (1999) destina a demostrar que la presencia de ciertas características fenoménicas es lógicamente suficiente para la presencia de ciertas características intencionales.

[16] Los pacientes con visión ciega, a causa de una lesión en el área occipital V1, "están ciegos en parte del campo visual, en el sentido de que no son conscientes de los estímulos presentados [en dicha parte]. A pesar de ello, son capaces de hacer discriminaciones y juicios acertados sobre los estímulos visuales presentados en dicha área ciega" (Viñuela Fernández, 2007: 237) cuando son forzados a emitir tales juicios, cabe añadir. Anotemos para terminar que, a pesar de que hemos presentado a Weiskrantz como el descubridor del fenómeno, un año antes de que acuñara el término (blindsight) Pöppel, Held & Frost (1973) habían ofrecido ya evidencia de su existencia.

[17] El típico paradigma experimental de *priming* pone de manifiesto el modo en que un estímulo no percibido conscienemente afecta a la ejecución de una tarea cualquiera, como clasificar estímulos pulsando botones o completar palabras partiendo de una serie de letras sin sentido. En un citadísimo estudio replicado en numerosas ocasiones, Neumann & Klotz (1994) presentaron evidencia empírica de efectos de *priming* aun cuando los *primes* (los señalados estímulos no percibidos conscientemente) se presentan enmascarados y se obtura su acceso a la conciencia.

blar de los fenómenos inconscientes que afectan a la conducta de los sujetos en cualquiera de ambas situaciones como intencionales dada su potencialidad consciente supondría que tanto la lesión en V1 como el escotoma en el campo visual del paciente con visión ciega debieran poder desaparecer, o que los sujetos experimentales de estudios del tipo del recientemente realizado por Marcos Malmierca (2014) pudieran diferenciar los estímulos que producen los contrastes entre sus respuestas ante las diversas situaciones experimentales –sobra añadir que, en un caso como en el otro, se trata de posibilidades imposibles, si se nos permite el oxímoron–. Hablar de potencialidad sólo tiene sentido cuando la misma puede de hecho actualizarse, cosa que no ha sucedido aún con la presunta potencialidad consciente de la actividad nerviosa que de los geniculados laterales parte errabunda hacia zonas circunstriadas sin dejarse ver previamente por V1 cuando a un paciente con visión ciega se le presentan estímulos visuales en el área ciega de su campo visual. Demos un paso más en esta dirección e imaginemos la siguiente viñeta. José opera a corazón abierto, *in extremis*, a Juan, que no sale de la operación. José abandona sudoroso el quirófano y le dice a la mujer de Juan: "ahora podemos estar seguros de que su creencia según la cual en el restaurante de la Torre Eiffel se sirve pan no era 'intrínsecamente' intencional, porque ahora sabemos que su posibilidad de actualizarse ha desaparecido y, con ella, su potencialidad". Ella, claro, se echa a llorar. Esta historieta debiera servir a dos propósitos. Por una parte, debiera hacer patente el modo en que la de la potencialidad consciente de un estado mental es, en última instancia, una noción malabarística e indefendible de la que resulta descabellado pretender extraer una demostración del pretendidamente necesario vínculo entre intencionalidad y fenomenalidad. Por otra, debiera ilustrar el sentido en que la concepción del inconsciente que Searle desarrolla no evita una versión neurofisiológica en absoluto plausible de la metáfora naïve de la mente como almacén de la que siente la tentación de reírse (Searle, 1992: 152 del original, 161 de la traducción). ¿Por qué? Porque, dado que cada estado intencional no actual no parece poder corresponder en su propuesta sino a estados neurofisiológicos concretos esperando su turno para ser iluminados por la luz de la conciencia, y dado que, como veíamos, Searle entiende el inconsciente como el conjunto de las actitudes proposicionales que un sujeto tiene mientras no es consciente de ellas, la neurofisiología de cada ser humano debiera ser infinita en todo momento, esto es, yo mismo debiera poseer un estado neurofisiológico concreto para mi creencia no actual según la cual en el restaurante de la Torre Eiffel se sirve pan, pero también para mi creencia no actual según la cual en el restaurante de la Torre Eiffel se sirve pan los mar-

tes, y también para mi creencia no actual según la cual, posiblemente, quepa elegir el tipo de pan que uno quiere que le sirvan un martes en el restaurante de la Torre Eiffel, y así sucesivamente.

Un par de comentarios finales acerca del chauvinismo inseparatista nos ofrecerán una imagen extrema de la posición de Searle, que ha llegado a afirmar que un estado mental puede contar como subjetivo y cualitativo exclusivamente cuando forma parte de "un campo consciente unificado" (Searle, 2007b: 170τ). Para contar como mental, de acuerdo con Searle, un *estado de cosas* tiene que formar parte de una vida mental unificada y compleja, pero lo cierto es que no tenemos ni idea de cómo ha de resultar ser una diminuta rana dorada panameña o incluso una avispa asiática gigante y gozar de formas simples de conciencia (como la primaria de Edelman o el *proto sí mismo* de Damasio, por ejemplo), y menos aun de qué grado debe alcanzar la complejidad de una vida mental para contar como un "campo consciente unificado". Por otra parte, partir del señalado modelo del inconsciente entendido como colección de actitudes proposicionales latentes, implica asimismo un compromiso chauvinista con una tesis según la cual lo fenoménico es necesario para la intencionalidad por su implicación con un aspecto de lo mental consistente en la capacidad de mantener actitudes hacia proposiciones, algo exclusivamente al alcance de criaturas lingüísticas como nosotros.

En resumidas cuentas, el único argumento de Searle en defensa de su tesis según la cual todo fenómeno mental se encuentra esencialmente vinculado con la conciencia (Searle, 1992: 20 del original, 34 de la traducción) parece ser su insistencia en la misma.

Listamos para terminar los principales entre los motivos que debieran mantenernos alerta frente al heroico encanto heterodoxo del inseparatismo.

1._Los inseparatistas se muestran inexcusablemente incapaces de sustentar sus tesis en evidencia empírica alguna y, lo que es peor, asimismo incapaces de sugerir vías que de algún modo permitieran acariciar semejante posibilidad.

2._Las disquisiciones (titubeamos en este punto entre las dos acepciones de esta voz en la vigésima segunda edición del diccionario de la lengua española de la RAE) con las que han rebasado la mera iteración de su dogma de la "esencialidad" de lo fenoménico son a menudo meros ejemplos destinados a ilustrarlo intuitivamente antes que a demostrar su justeza o, en el mejor de los casos, remedos de argumento en cuyas premisas se introducen de contrabando las que habrían de ser sus conclusiones.

3._Como tercer motivo contaría la más que probable gratuidad de las asunciones ontológicas que vinculamos con nuestra economía léxica. Bien cabe que «intencionalidad» y «fenomenalidad» no pasen de ser sustantivos utilizados para aludir a conjuntos un tanto heterogéneos de fenómenos. No obstante, la práctica habitual, común a separatistas e inseparatistas, consiste en tratar dichas nociones como si de conceptos netamente unívocos que refirieran de forma inequívoca a fenómenos completamente discretos se tratara, como si cada una de dichas nociones portara, por así decir, una flecha ontológica libre de toda duda, una flecha inhesitablemente capaz de acertar en la diana de su singular referente. Adicionalmente, si dichas nociones refieren de forma unívoca e inequívoca, y si los defensores de las soluciones separatista e inseparatista están en lo cierto, habrán en cualquier caso las mismas de hacerlo a sendas características o propiedades de diferentes clases de fenómenos, sucesos u objetos. Pero, para decirlo en dos palabras, la transparencia y la liquidez del agua son dos características diferentes de una y la misma clase de objeto, pero a pocos se les ocurriría preguntar si una explica la otra o viceversa, porque a pocos se les escapa que quizá el estatuto ontológico del agua resulte menos discutible que el de nuestro descriptor "transparente". Puede que una buena idea consista en olvidar estos supuestos órdenes de dependencia, centrarnos en aclarar enredos conceptuales relevantes –en lugar de crear nuevos enredos sin visos de ir a rentar en la crítica y desarrollo de marcos teóricos vinculados a programas experimentales o en la interpretación y discusión de los resultados obtenidos dentro de cualquiera de tales marcos–, refinar nuestros marcos teóricos y continuar recopilando y articulando datos en psicología, etología y neurociencias.

4._El inseparatismo, tal y como la propuesta de Searle pone elocuentemente de relieve, conduce a una concepción dualista de la realidad según la cual existen, por una parte, organismos conscientes complejos y, por otra, ciegos dispositivos mecánicos. El inseparatismo implica así una dicotomía de todo o nada entre organismos conscientes complejos y un mundo objetivo completamente ciego (vid. Shani, 2007). Existen, en esta línea, razones intuitivamente sólidas para poner en tela de juicio la idea según la cual la búsqueda de criterios estrictos de los que servirnos para delimitar de forma limpia y rígida los fenómenos mentales distinguiéndolos de sus precursores evolutivos vaya a ser una empresa exitosa, razones que tienen en último término que ver con el hecho de que resulta extremadamente difícil encontrar en el árbol filogenético el lugar exacto en que la primera pluma se diferenció de la última escama o el momento exacto en que rompió el cascarón el primer ave o el último reptil. "Antecedentes esencialmente hete-

rogéneos" no significa nada en biología, entre otras cosas porque "esencial" no significa nada en biología, al menos no desde el 24 noviembre de 1859. Esta última consideración nos lleva al último problema que consideraremos.

5._El inseperatismo conduce ineluctablemente a una última dicotomía, dado que nos obliga a elegir entre emergencia súbita o panpsiquismo. En pocas palabras, si todas las características definitorias de lo mental han de venir juntas, sólo le quedan a lo mental dos opciones: o surgir de lo físico abruptamente, de repente y sin precedentes –o con "antecedentes *esencialmente* heterogéneos"–, o ser inherente a ello. Venimos nuevamente a parar así en el catastrofismo cuvieriano como única alternativa a una concepción según la cual cada tipo de leptón tiene su particular clase de deseos, creencias, ansiedades y picores.

## 5._Conclusión

Como hemos visto, las opciones que monopolizan el mercado filosófico no contribuyen sino a enmarañar la ya de por sí fragosa maraña tradicional y no problematizada de categorías, asunciones y contraposiciones enzarzándose en discusiones cuyos visos de ir a servir para allanar el camino hacia una solución al problema de la conciencia se muestran escasamente halagüeños. Los bandos enfrentados en dichas discusiones comparten, como hemos podido comprobar a lo largo de este capítulo, una serie de asunciones difícilmente justificables. Desafiar la principal de entre las mismas, la de que los términos en los que hoy se plantea el debate son los términos en los que debe plantearse, ha sido el objetivo del capítulo a cuyos últimos párrafos se enfrenta el lector. Esta asunción deriva en la idea de que la solución al problema de la conciencia depende –de un modo y por unos motivos que nadie ha logrado explicitar– de la especulación acerca de la relación entre dos abstracciones que aparecen impresas en una de cada dos carillas publicadas en el área de los *Consciousness Studies*: la fenomenalidad y la intencionalidad. Esta idea se apoya, a su vez, en una serie de supuestos cuya labilidad hemos intentado asimismo poner de manifiesto. Entre estos supuestos, el de la unívoca referencialidad de dichas abstracciones ha sido al que más atención hemos prestado.

Separatistas e inseparatistas comparten la referida serie de asunciones y supuestos. Unos y otros tratan de contribuir al progreso de los *Consciousness Studies* avanzando y discutiendo definiciones de los términos que, entienden, han de fundamentar la actividad investigadora en el área y olvidando que, como indicamos, las definiciones suelen

llegar tarde en el desarrollo de cualquier área de investigación, ganando habitualmente contenido y solidez conforme el área adquiere madurez. Este especulativo trajín lexicográfico ha venido configurando el debate contemporáneo, cargándolo de inercias metodológicas, confusiones conceptuales y enredos teóricos tan desorientadores como infructuosos.

Hemos podido comprobar, por otra parte, que especular es gratis. Si se nos permite dedicar a tal tarea un par de renglones, quisiéramos añadir para cerrar nuestra crítica de las señaladas asunciones que muy probablemente las mismas procedan, en último término, de una tradición intelectualista que ha venido definiendo, desde Descartes, a la mente en términos cognitivos, intelectuales, en términos de cálculo y representación, una tradición que vemos desembocar hoy en la idea de que explicar la conciencia es algo que sólo puede conseguirse especulando acerca de su relación con eso que, latamente, podemos denominar "mente cognitiva". Según la apuntada tradición, trafagar con "información" es lo que define a lo mental y, por lo tanto, a esa actividad habrá que atender para explicar todos y cada uno de los fenómenos mentales. La metáfora computacional de la manipulación de información ha ofrecido la ocasión de avanzar esclarecedoras hipótesis internalistas acerca del funcionamiento de la memoria, la percepción o el lenguaje. No obstante, tal y como Mosterín (2003b) hace notar, una cosa son las explicaciones y teorías científicas, y otra las metáforas que, mediante conjeturas tentativas formuladas echando mano de conceptos remisos permiten dar los primeros pasos hacia la comprensión de fenómenos cuya naturaleza y forma de funcionamiento nos son aún inaccesibles. Sea como fuere, en esta metáfora de la manipulación de información y en esta tradición intelectualista, desafiada hoy desde el enactivismo (vid, v. g., Noë, 2009: cap. 5), se apoya el reduccionismo representacional de los separatistas, pero tampoco quienes, como los inseparatistas, han pretendido dar la espalda a dicha tradición y dicha metáfora han sabido dejar atrás la idea de que la especulación acerca de la relación entre lo "fenoménico" y lo "intencional" podrá, de un modo y por unos motivos que nadie ha logrado hacer explícitos, abrir el camino hacia la solución del problema de la conciencia. Son, por otra parte, muchas y muy interesantes las capacidades mentales que, en una u otra medida, pueden plegarse a esta parva metáfora de la manipulación de información. Tales capacidades conforman esa "mente cognitiva", ese segmento de lo mental encargado de captar información, almacenarla y recuperarla, elaborar partiendo de ella planes, tomar decisiones y resolver problemas de álgebra lineal. Pero entreverada con esta mente cognitiva encontramos una mente que cabe denominar motivacional, y aun otra que

cabe denominar afectiva. Dado que la tradición dirigió inicialmente su interés a la mente cognitiva, diseñando gran cantidad de herramientas teóricas y metodológicas a fin de dar cuenta de la misma, muchos siguen hoy tratando de explicar cualquier otro aspecto de lo mental haciendo uso exclusivo de esas herramientas de cuño intelectualista, un proyecto carente de toda justificación más allá de la inercia, pues parece a todas luces evidente que "no hace falta mucho esfuerzo intelectual para experimentar dolor, miedo o hambre" (Stamp Dawkins, 2000: 883τ). De este modo, difícilmente podrán abrirnos paso hacia la solución del problema de la conciencia las discusiones que mantienen entretenida a la práctica totalidad de los teóricos implicados en el área de los *Consciousness Studies* en un redundante ejercicio de dar nuevas vueltas en torno al poste intelectualista. Pero debe haber alguna forma de hacerlo, alguna forma de propiciar un cauteloso avance hacia dicha solución. ¿Cuál? A responder a esta pregunta dedicamos el siguiente capítulo.

# CAPÍTULO 9

## REPLANTEAMIENTO DEL NÚCLEO DEL DEBATE. HACIA "UNA" EXPLICACIÓN DE LA CONCIENCIA

### 1._Qué significaría resolver el problema de la conciencia

En los capítulos cuarto y quinto nos acercamos al estado del arte de las teorías de la conciencia: el cuarto nos brindó la ocasión de comentar las ontológicas mientras el quinto nos ofreció la de hacer lo propio con las explicativas. ¿A qué punto hemos llegado ahora? Respecto de las propuestas ontológicas no diremos nada, al menos no de forma directa. Entendemos que el problema de la conciencia es el que tratan de resolver las propuestas explicativas y que el núcleo del mismo se halla así en las dificultades que venimos encontrando a la hora de esclarecer las causas de su existencia y los mecanismos de su funcionamiento antes que en esas otras que se nos presentan cuando intentamos formular una definición de lo que es. Como es bien sabido, "nadie tiene ni la más remota idea de qué es la conciencia" (Fodor, 2004: 31τ) y, al tiempo, "todo el mundo sabe lo que es la conciencia" (Stout, 1899: 7τ[1]; Edelman & Tononi, 2000: 3 del original, 15 de la traducción; Tononi, 2008: 216τ; 2012b: 293τ), de forma que quizá la mejor solución al problema ontológico, el de tratar de especificar qué es la conciencia, consista en una mezcla de silencio y paciencia. El problema explicativo viene primero. Cuando comprendamos por qué existe y cómo funciona la conciencia nos encontraremos en mejor situación para discutir acerca de qué es y cuál debe ser el lugar que ha de ocupar en nuestro esquema de la realidad. Tratar de definir qué es la conciencia nos suena, pues,

[1] Es interesante hacer notar el modo en que este fragmento de este manual clásico es citado poco después de forma crítica y sin mención explícita de la procedencia del texto en otro manual clásico, los *Psychological Principles*, de James Ward (Ward, 1918: 21).

precipitado: como señalamos al principio del segundo capítulo, la ciencia acaba por ofrecer definiciones más que seguirse de ellas. Adicionalmente, no resulta extemporáneo hacer notar que justificar la existencia de la ontología entendida como disciplina, con unos métodos y resultados definidos, es un proyecto que nadie ha llevado a su término. Nuestras prácticas epistémicas penetran la ontología configurando nuestra comprensión de la misma, lo cual no viene a significar que ellas la determinen, sino que, sencillamente, determinan nuestro acceso a la misma. Nuestra ontología es, por lo tanto, producto de dichas prácticas. Después de ellas, el trabajo del ontólogo, si cabe hablar de algo semejante, será el de trazar nexos significativos dentro de la enmarañada red de resultados y teorías científicas. Gran cantidad de disciplinas y marcos teóricos son necesarios para explicar cumplidamente la locomoción o la reproducción. Lo mismo sucede con lo mental. Cada día aparecen publicaciones sobre la ontología de lo mental. ¿A nadie preocupan las de la locomoción o la reproducción? La ontología de lo mental no supone un problema en sí misma, es decir, no supone ningún problema diferente del de nuestro acceso epistémico a lo mental, del de nuestro conocimiento científico de lo mental. Esto parecen comprenderlo todos los ontólogos de la locomoción y la reproducción, pero apenas ninguno de los ontólogos de lo mental que retocan en estos momentos el borrador de su último artículo. El problema explicativo no es sólo la base para abordar el ontológico: es *el* problema de la conciencia. Como veremos a lo largo de este capítulo, su solución implica la del resto de los problemas de la conciencia a los que aludiéramos en la primera parte, pues no podemos explicar la conciencia sin describirla apropiadamente (problema descriptivo), como tampoco podemos hacerlo sin comprender su función (problema funcional), una comprensión que presentará al problema causal como un mero enredo conceptual y que no podrá deslindarse de la indagación de la génesis de la experiencia que requiere el que denominamos problema del punto de ebullición.

Sobrepuja el problema explicativo al resto, que, por así decir, posponemos a la elucidación de aquél. Pero, con todo y con esto, incidamos: ¿a qué punto hemos llegado? ¿Dónde nos encontramos tras nuestra crítica de las soluciones separatista e inseparatista? Nuestra crítica del inseparatismo no anula ninguna opción explicativa, porque tampoco comporta el mismo ninguna, sino sólo vagas intuiciones metafísicas y en ningún caso nada parecido a orientaciones heurísticas, metodológicas o epistemológicas. Aun así, desvelar qué clase de discusiones han venido sirviendo, exclusivamente, para desviar la atención hacia problemas espurios y sembrar el debate contemporáneo de controversias

que en ningún sentido pueden contribuir a avanzar hacia una solución del problema de la conciencia será de utilidad en el mismo sentido en que abandonar hábitos alcohólicos lo es de cara a convertirse en un deportista de élite: no basta la omisión, pero es necesaria para centrarse en lo que ha de seguirla. Por su parte, nuestra crítica del separatismo parece dejar fuera de juego todos los intentos explicativos cognitivos y representacionales, con lo cual optarían a fungir como explicación de la conciencia sólo algunas de las teorías neurobiológicas disponibles en el mercado contemporáneo. No obstante, quisiéramos hacer un doble llamamiento a la prudencia. En primer lugar, cabría objetar nuestro ataque a la solución separatista argumentando que la noción de representación que con el mismo poníamos en tela de juicio es una noción injustificadamente inflada y que, así, nuestra crítica de las teorías de la representación sería más bien una crítica de las teorías de la intencionalidad. Introdujimos la distinción que cabría trazar entre representación e intencionalidad en el capítulo tercero al señalar que la noción de intencionalidad con la que se trafaga en la literatura contemporánea parece implicar más que la mera relación representacional (el mero *aboutness* o ser-acerca-de) al incluir aspectos netamente semánticos. Sin embargo, y aunque pudiera parecer que hay espacio para una réplica de este tipo, lo cierto es que una teoría de la representación que no dé cabida a la normatividad y la aspectualidad, es decir, una teoría de la representación que no maneje un concepto de representación según el cual las representaciones puedan, de un modo u otro, ser falsas y representar sus objetos bajo unos aspectos en lugar de bajo otros, una teoría tal, pues, no sabemos muy bien de qué sería una teoría. Así, defender que las teorías de la representación hasta ahora ensayadas son teorías infladas de chauvinismo semántico que no hacen justicia a la naturaleza de la representación al introducir de contrabando en su consideración rasgos semanticistas que exceden el mero y elemental *aboutnes* no serviría para eludir la base de nuestra crítica al inseparatismo fundada en la ausencia de una teoría de la representación consensuada y en la implausibilidad biológica del proyecto de elevar la explicación de todas las formas de interacción con el entorno concebidas como representacionales sobre la sola base de especulativos análisis apriorísticos de esta o aquella noción de representación, sino para apuntar a la necesidad no ya de poner un nuevo parche a las teorías de la representación hasta la fecha elaboradas, sino a la de desechar las últimas cinco décadas de elucubración en torno al problema de la representación para reformularlo en términos más básicos. Pero, por una parte, lo cierto es que nadie sabe cuáles habrían de ser esos términos más básicos porque, insistamos, el concepto de representación según el cual una representación no puede ser

en algún sentido falsa pero puede representar sin hacerlo de forma aspectual ni parece ser del que vienen tratando los científicos cognitivos ni ningún otro en el que nos quepa pensar de forma coherente, con lo cual la distinción trazada en el capítulo tercero entre el mero y elemental ser-acerca-de y la intencionalidad propiamente dicha ha de ser considerada con toda prudencia, pues hablar de chauvinismo semántico en relación a la aspectualidad y la normatividad parece, por lo antedicho y cuando menos, un tanto problemático –harina de otro costal han de considerarse la idea de la individuación de grano fino y la de la racionalidad mínima–. Además, por otra parte, cualquier reformulación del problema de la representación, por básicos que sean los términos en que se proyecte, seguirá presentándonos la clase de lo que denominamos representacional como homogénea al punto de poder explicarse desde una sola y concretísima teoría fundamental. Pero, ¿han de resultar exhaustivamente explicables todos los fenómenos biológicos que denominaríamos representacionales dentro de un mismo nivel explicativo y de acuerdo a unos mismos principios teóricos, destilados, por añadidura, en el alambique de la especulación a priori? Nada lo asegura, y la frondosidad del árbol filogenético, así como la pluralidad filogenética y fisiológica de las economías de interacción cognitiva con el entorno que puebla sus diferentes ramas sugieren que un puñado de especulativos principios teóricos apriorísticos es menos de lo necesario para dar cuenta de tan abigarrado marco. ¿Existe algún motivo de principio que excluya la posibilidad de que a los fenómenos biológicos que incluimos en la clase convencional de lo representacional subyazcan, en pocas palabras, diferentes clases de fisiologías y filogenias explicables en distintos niveles desde diferentes marcos teóricos? Inclinándonos a responder negativamente, y a la vista de las dificultades a que conduce la réplica fundamentalista de la reformulación del problema de la representación en términos básicos, entendemos que esta primera llamada a la prudencia podría dejarse de lado, pero, en cualquier caso, no somos futurólogos: quizá mañana se logre una singular teoría fundamental, básica y omniexplicativa de la representación capaz de hacer las veces de resorte único y último del que derivar y sobre el que asentar la explicación de toda forma de transacción mental entre organismos y entornos, una teoría ensayada por vías más sutiles y cabales que las exploradas hasta la fecha y válida para todo lo ancho y largo del árbol filogenético.

En segundo lugar, nuestro llamamiento a la prudencia cobra sentido desde una interpretación de nuestra crítica de la solución separatista en la cual no aparece ésta como un rechazo esencialista de los resultados empíricos, paradigmas experimentales e, incluso, marcos teóricos gestados dentro de la tradición cognitivista. Procedamos analógica-

mente. La hegemonía conductista tuvo su momento; no obstante, una enorme cantidad de resultados empíricos, herramientas teóricas y paradigmas experimentales gestados dentro de esta tradición siguen con nosotros, como hace evidente una rápida consulta de cualquier manual de psicología del aprendizaje o un superficial vistazo al uso que de los mismos hacen la práctica totalidad de las ramas de las neurociencias. Una explicación de la conciencia, si es que este uso singular tiene algún sentido, habrá de ser algo tan extraordinariamente rico y plural que cuesta imaginar el sentido en que a algunos parece antojárseles pertinente ya no la posibilidad, sino incluso la necesidad de su elaboración de espaldas a casi tres cuartos de siglo de resultados empíricos, paradigmas experimentales y marcos teóricos cognitivistas en psicología de la percepción, psicología de la memoria, psicología del pensamiento o psicología de la atención. El eclecticismo es considerado a menudo como la debilidad del que no se decide, como la resolución del que no se resuelve. Siempre resulta más vistoso un aspaviento seguido de la exhortación: "todo aquello se acabó, ¡empecemos de nuevo!". Se busca hoy la tierra ignota de la explicación de la conciencia y pocos parecen apetecer sino propuestas radicales. Una lástima, porque "puede que el radicalismo sea bueno en política, pero es malo para la ciencia" (Prinz, 2009: 419τ), y porque bien cabe que, de este modo, el borrón y cuenta nueva sea lo menos indicado no sólo por lo que a los resultados empíricos y los paradigmas experimentales se refiere, sino asimismo por lo que a los marcos teóricos respecta, pues en ciencias "con frecuencia nos encontramos con que diversas teorías tienen ventajas distintas y diferentes ámbitos de aplicación o validez" (Mosterín, 1984/2000: 312). La explicación de la conciencia tendrá muchas capas, y a nadie debiera extrañar que algunas de las mismas se muestren antes dóciles ante determinadas herramientas teóricas de raigambre cognitivista que ante cualesquiera otras. Una tal explicación habrá de acomodar gran cantidad de fenómenos descubiertos y explicados hasta hoy dentro de paradigmas experimentales y marcos teóricos cognitivistas, y lo cierto es que no hay motivos de principio contra la posibilidad de que determinados segmentos de determinadas interpretaciones cognitivistas de dichos fenómenos puedan ser de utilidad en el trazado de nexos explicativos entre o dentro de algunas de las señaladas capas. "La comprensión de la mente (…), en todos sus niveles de organización, incluidos el de la conciencia y la cultura, va a requerir ir más allá del procesamiento de información como la forma principal (…) de acercarse a su estudio, sin que ello suponga desacreditar la indudable aportación empírica y teórica de [este paradigma]" (Pozo, 2001: 97).

Que la explicación de la conciencia haya de ser algo plural, algo con muchas capas, parece ser una de esas obviedades que la mayoría de la bibliografía contemporánea desatiende. La pluralidad de disciplinas y marcos teóricos requeridos para la explicación de un fenómeno cualquiera es algo que depende de la complejidad del fenómeno que sea el caso. Un ejemplo simple. Tras una mutación en un determinado locus del cromosoma circular de una cepa de determinada especie de bacteria sobrevenida por exposición a la luz solar, una nueva forma de un determinado alelo se impone tras unas cuantas generaciones y los ejemplares de dicha cepa se benefician de una capacidad ampliada para aprovechar los nutrientes disponibles en su entorno. La explicación de este fenómeno, descrito en cuatro renglones, requiere de la combinación de marcos teóricos provenientes tanto de la física clásica como de la no clásica[2] (v. g., óptica, física molecular, física atómica y nuclear, química física, etc.), la bioquímica o la genética de poblaciones. No hay un único marco teórico del que quepa derivar *la* explicación de este fenómeno. Las explicaciones científicas son algo singular sólo en abstracto, sólo en la mente de los filósofos. En la práctica, la explicación de cualquier fenómeno posee siempre múltiples capas. Pero dado que existen marcos teóricos que pueden entenderse como más abarcadores y, por tanto, como capaces en principio de integrar en su seno la explicación de fenómenos que caerían dentro de otros más particulares, y dado que existe una extendida concepción del progreso científico como tendente a la simplificación y la generalización, viene vaticinándose el día en que alguien dará a luz una macroteoría capaz de explicarlo todo: la teoría del todo. Hasta que ese día llegue, seguiremos viviendo en un universo analizable en múltiples niveles y parcialmente explicable gracias la indescriptiblemente plural plétora de herramientas teóricas hasta la fecha diseñadas. Que la explicación de la conciencia haya de plantearse como una empresa plural e interdisciplinaria, articuladora de múltiples niveles, tiene que ver con el hecho de que el desacierto de concebir a la conciencia como un fenómeno homogéneo, discreto y unitario reside en que todo apunta en la dirección contraria invitando a depositar tantas esperanzas en el proyecto de elaborar *una* explicación de la conciencia como en el desarrollar una teoría unificada de la mente. No existe hoy algo así como *una* explicación de la mente –y cabe dudar que vaya a hacerlo mañana–, sino una enorme cantidad de grupos de investigación trabajando en diferentes problemas dentro de diversos marcos metodológicos y teóricos. Sin embargo, hay abierta actualmente una guerra entre diferentes

[2] Actualmente existe unanimidad acerca del rol de eventos cuánticos en la génesis de mutaciones (vid., v. g., Ruvinsky, 2010: cap. 2).

paradigmas en ciencias cognitivas que, curiosamente, parece entretener más a los filósofos que a los investigadores que, en cierta medida ajenos a la misma, siguen trabajando dentro de sus respectivos marcos metodológicos y teóricos. Las batallas más encarnizadas de esta guerra interparadigmática tienen lugar dentro de las lindes de los *Consciousness Studies*: gran cantidad de teóricos, filósofos, principalmente, parecen estar extremadamente interesados en demostrar que la conciencia habrá de explicarse, exclusivamente, desde esta o aquella forma de cognitivismo o enactivismo. Con todo, puede que esta guerra sirva para poco más allá del emborronado de cuartillas. La conciencia se encuentra relacionada con prácticamente cada proceso mental. El que crea que ha de existir *una* teoría unificada de lo mental tiene sobrados motivos para presentar batalla, muy probablemente más de los que cabría encontrar a la base de dicha creencia. Por su parte, el que crea atisbar en la historia de las ciencias cognitivas una ineludible tendencia a abordar y acomodar fenómenos y procesos mentales particulares dentro de marcos metodológicos y teóricos particulares, encontrará absurda la señalada actitud beligerante. ¿Han de plegarse absolutamente todos los fenómenos que denominamos mentales a dos o tres paradigmas experimentales y una o dos generalizaciones? ¿Es razonable esperar una *teoría cognitiva del todo* capaz de integrar en un conspicuo hatajo de enunciados legaliformes cada uno de los fenómenos y procesos estudiados por la psicología de la atención, la de la memoria, la ciencia del sueño, la psicofarmacología, la neuroetología, la psicología comparada, la fisiología de la conducta, la neurolingüística, etc., etc.? Puede que unos cuantos físicos teóricos encuentren razonable prolongar, a través de once dimensiones hoy y mañana quizá veintiséis, su carrera hacia la meta de un puñado de leyes omniexplicativas, pero probablemente el universo de lo mental sea excesivamente complejo para que resulte hacedera la pretensión de confinarlo dentro de la férula de un solo paradigma omnipotente: bien cabe esperar que el enactivismo logre refinar los paradigmas experimentales e iluminar con solvencia buena parte de los principios explicativos de la psicología de la percepción mientras a determinada forma de simbolismo o conexionismo le sea dable hacer lo propio con los de la psicología del pensamiento. Como a menudo se ha defendido respecto de la necesidad de integrar herramientas teóricas conductistas y cognitivistas de cara a dar adecuada cuenta del modo en que se produce el control de la conducta (vid., v. g., Toates, 1998), es posible que cada uno de los bandos de esta guerra interparadigmática haya capturado ya parte de la verdad, pero también lo es que a ninguno le quepa capturarla acabadamente en solitario. En resumidas cuentas, dada la pluralidad y complejidad del fenómeno a explicar –un fenó-

meno que aparece en el árbol filogenético, con toda seguridad, en múltiples formas y grados, e implicado en prácticamente cada aspecto de la vida mental de los organismos en los que tiene lugar–, escasa admiración debiera suscitar que su explicación requiriera de una aproximación plural y compleja, capaz de integrar diversas disciplinas y marcos teóricos y, así, de soslayar las implicaciones esencialistas de la endémica y estéril batalla interparadigmática que se libra hoy en la arena de las ciencias cognitivas. Quizá sea demasiado pronto para una teoría cognitiva del todo, y puede que siga siéndolo aún cuando la naturaleza de lo mental sea exhaustivamente desentrañada. El límite a la cantidad de estrategias, procedimientos, instrumentos, marcos teóricos, herramientas heurísticas y tipos de datos potencialmente útiles será aquí, como en cualquier área de investigación, extremadamente difícil de atisbar. Tanto el carácter del fenómeno a explicar como el actual estado de nuestro conocimiento científico del mismo invitan, pues, antes que a lanzarse a la romántica caza de una teoría unificada y fundamental, a seguir acumulando datos empíricos y perfilando modelos teóricos ceñidos a aspectos concretos del fenómeno en cuestión –y no pocos se inclinan a pensar que una vez elucidado convenientemente cada uno de ellos no quedará nada más que explicar.

> La conducta y la [mente] son actividades de sistemas que atraviesan diversos niveles de lo real (…), niveles desde el físico hasta el social. Por tanto, ninguna ciencia que se ocupe de un solo nivel las explicará. Siempre que el objeto de estudio es un sistema con múltiples niveles, lo único promisorio es un enfoque multidisciplinar –un enfoque que cubra todos los niveles que intervienen. En estos casos la obstinación en el reduccionismo está condenada al fracaso puesto que insiste *ab initio* en procedimientos que no pueden ser puestos en práctica por falta de hipótesis acerca de las relaciones entre niveles (…). Presionar para que en estos casos se efectúe una reducción es quijotesco: no es una estrategia promisoria de investigación. La única que en estos casos puede alcanzar el éxito es una estrategia oportunista (*catch-as-catch-can*) sugerida por el sistemismo y por una concepción del mundo como compuesto de múltiples niveles. (Bunge, 1980/1988: 215 del original, 225-228 de la traducción).

Eclécticos pragmatismos a parte, y matizando nuestro llamamiento a la prudencia, la tesis representacionalista dura, esto es, en la versión que suelen defender los filósofos, no se presenta como una opción razonable, dado que la misma, como hemos visto, aparece envuelta en las brumas de una futurología de la ciencia completamente inespecífica y dogmática. ¿Qué futurología? ¿Qué dogma? La siguiente cita debiera servir no sólo para resumir nuestra crítica de la solución separatista, sino asimismo para responder a las preguntas que acabamos de formular. “A working premise behind the Representational Thesis is that a better understanding of *the mind* is not to be obtained by knowledge –no matter how detailed and precise– of the biological machinery by means

of which the mind does its job" (Drestke, 1995: xiv).[3] El dogma: no importa cuán lejos le quepa a la biología llegar, pues sabemos, desde hoy, a priori, que nunca será suficiente. La futurología: también desde hoy y a priori sabemos que todos los misterios de lo mental habrán de diluirse en una –a día de hoy enteramente misteriosa– teoría naturalista de la representación acerca de la cual, por lo pronto, cuanto cabe avanzar es que nadie sabe qué hay que entender por «teoría» ni qué por «naturalista» ni qué por «representación» cuando se habla de una futurible teoría naturalista de la representación. Contra esta tesis representacionalista, defenderemos la necesidad de tomarse en serio la idea según la cual la conciencia es un fenómeno biológico y, por tanto, de cuño biológico ha de ser su explicación –si es que acaso, insistamos, tiene en este punto sentido hablar en singular–. Pero dicha necesidad puede plantearse desde dos perspectivas. Por una parte, investigando la fisiología de la conciencia desde un enfoque neurobiológico tradicional, en la línea de las teorías neurobiológicas comentadas en la primera parte. Por otra parte, apelando al carácter extendido y relacional de lo mental, cuyo locus se ubica, ubicuo, en la interacción del sistema nervioso con el resto del organismo y de éste con su medio. Los teóricos que así han venido haciéndolo, a pesar de estar enteramente de acuerdo con Eric Kandel cuando afirma que "el principio que subyace a la nueva ciencia de la mente es que *todos* los procesos mentales son biológicos" (Kandel, 2006: 336 del original; 390 de la traducción),[4] entienden que la noción de biología que utiliza resulta excesivamente estrecha, dado que, en concordancia con el primer principio postulado en Kandel (1998), el Nobel de origen austriaco cierra la frase que acabamos de citar indicando que los procesos biológicos relevantes para el investigador de lo mental se desarrollan dentro del cráneo.[5] Desde esta segunda perspectiva, la neurobiología de la conciencia no ha dado con la solución al problema de la conciencia precisamente porque ha venido buscándola en el interior de nuestros cráneos, como si los fenómenos mentales tuvieran lugar ahí dentro, como si pudieran localizarse espacialmente, como si pasaran sus vidas confinados en sus cámaras óseas. Muy al contrario, según estos teóricos, tales fenómenos no ocurren en nuestro sistema nervioso, sino en la interacción del organismo al completo con su medio. Ciertamente, esto mismo sucede con absolutamente todos rasgos de los seres vivos: ellos son, sin excepción y tanto desde el punto de vista filogenético como

[3] Cursivas en el original.
[4] Cursivas en el original.
[5] John Allan Hobson ha expresado la misma idea de la forma más concisa al afirmar que el cerebro es consciente (Hobson, 2001: 6 del original, 23 de la traducción; 17 del original, 34 de la traducción).

desde el ontogenético, producto de la interacción con el medio. Sin embargo, lo mental parece serlo en grado superlativo: lo mental es el órgano de la relación.

Al igual que los esteroides sexuales contribuyen durante la embriogénesis a conformarnos ontogenéticamente, nuestra relación con el medio nos configura en un sentido filogenético: las especies no evolucionan danzando sobre el vacío, sino inmersas en la trama de relaciones que definen sus nichos ecológicos. En este sentido, esa relación tiene un carácter constitutivo. Pero, además, al igual que los esteroides sexuales contribuyen en la activación de la conducta sexual adulta de forma dinámica, esto es, viéndose implicados no en la conformación de una estructura, sino en su puesta en funcionamiento en un determinado sentido, asimismo es nuestra relación con el medio la dinámica condición de posibilidad de nuestro flujo de actividad. En este sentido, esa relación tiene un carácter no sólo constitutivo, sino también dinámico, actual y muy real: hace de nosotros eso que somos en un sentido filogenético, como apunábamos, pero lo hace asimismo a cada instante.

Cualquiera de nuestros rasgos es, pues, en cualquier sentido relevante concebible, fruto y condición de nuestra forma de relacionarnos con nuestros medios. Pero ninguna característica de los organismos queda hasta tal punto definida por su relación con el medio como sus características mentales. Tanto es así que algunos vienen abriendo camino a la ampliación de nuestra noción de lo mental para que ella dé cabida a partes "no-neuronales" del cuerpo. Nada es en un animal a tal punto relacional como lo mental. Antes de su emergencia evolutiva, desde luego, los organismos se relacionaban eficazmente con sus medios: se desplazaban, se alimentaban, se reproducían, etc. Pero la relación con el medio desde el punto de vista del desplazamiento, la alimentación o la reproducción difiere de la forma de relación a que obedeciera la génesis de lo mental, porque la misma vino a responder, precisamente, a la necesidad de ampliar, refinar y flexibilizar los repertorios de interacción de los organismos con sus medios. Así, por ejemplo, desde la perspectiva de la forma de interacción implicada por la alimentación, ante un determinado ser vivo caben para otro, exactamente, dos opciones: me lo como, no me lo como. Lo mental abre la vía de innúmeras posibilidades: "ahora esconderé su cadáver, más tarde vendré a comérmelo", "temo que me devore: he de ocultarme", etc. Todos nuestros rasgos dependen filogenética, ontogenética y actual y dinámicamente de nuestra relación con nuestros medios, pero esa relación es para lo mental no sólo un determinante, sino asimismo una tarea, el material sobre el que trabaja al tiempo que el producto de su trabajo.

Estamos enteramente de acuerdo con los defensores de esta segunda perspectiva en que el substrato de los fenómenos mentales excede la actividad neurofisiológica, y en que si resultaba difícil entender el modo en que los postulados separatistas podían contribuir a vertebrar una explicación de la conciencia en sentido estrecho, ahora, es decir, en este contexto biológico más amplio, esa dificultad se torna imposibilidad. Sin embargo, a diferencia de ellos, no vemos cuál será el provecho de rechazar la supuesta ortodoxia *mainstream* de la neurobiología de la conciencia concebida como disciplina encargada de trabajar con hipótesis acerca del funcionamiento del sistema nervioso de los organismos conscientes, un funcionamiento, obviamente, contextual, íntimamente ligado al del resto del organismo en interacción con su medio. Tampoco alcanzamos a imaginar el sentido en que enarbolar esta o aquella bandera ayudará a avanzar hacia una ciencia de la experiencia consciente. A diferencia de nosotros, reformistas radicales como Hutto & Myin (2013) o Noë (2009), ven ese provecho y alcanzan a imaginar ese sentido, pues parecen pretender que algo fundamental marcha mal en las ciencias cognitivas en general y en la neurobiología de la conciencia en particular, de tal modo que explicar la conciencia será un proyecto irrealizable hasta el momento en que el enactivismo salga victorioso de la guerra interparadigmática relegando al resto de marcos teóricos al estatus de inútiles reliquias de la historia ciencia. Aquí, donde muchos han venido buscando y creyendo encontrar razones para presentar batalla en la guerra de los paradigmas, encontramos nosotros motivos para trascenderla, porque "desde que el objeto se muestra como un complejo de relaciones es necesario aprehenderlo por métodos múltiples" (Bachelard, 1934: 18 de la traducción). Aquí, por otra parte, se suman, dos dogmas contemporáneos interrelacionados: el de la pertinencia de la guerra interparadigmática, tras la cual se alzará triunfal *el* paradigma, y el de la unidimensionalidad de la explicación de la conciencia, según el cual la conciencia es *un* fenómeno con *una* explicación. Estos reformistas radicales, como no podía ser de otro modo, no dejan de admitir el papel decisivo que el sistema nervioso juega en nuestras vidas mentales (vid., v. g., Noë, 2009: 80 del original, 108 de la traducción), pero suponen que seguir investigando el modo en que lo hace e iluminando su fisiología no llevará de por sí a ninguna parte y desatienden con ello el hecho de que ésta es aún una tarea inconclusa. Cierto: una vez desentrañada la neurofisiología de la conciencia habrá de ser la misma incardinada en el contexto más amplio de la interacción entre sistemas nerviosos, cuerpos y entornos, y es de hecho más que probable que desentrañarla implique ya atención a esa incardinación, pero esto nada dice acerca de la unidimensionalidad de la explicación de

la conciencia ni acerca de la inadecuación de la investigación de su neurofisiología en el marco de paradigmas que tanto da que sean aquí o allá tachados de ortodoxos o tradicionales mientras alcancen a ofrecer resultados manifiestos. Cierto también, por otra parte, que la neurofisiología de la conciencia difícilmente puede investigarse en abstracto, sin tomar en consideración su relación con la dinámica de su interacción con cuerpos y entornos, pero cierto asimismo que dicha fisiología es aún desconocida y que nada asegura que ampliar su contextualización experimental servirá para algo más allá del aumento exponencial de la complejidad y la cantidad de variables implicadas en su investigación.

## 2._Qué significa explicar en biología

Dado que las características mentales que nos interesa explicar forman parte de la clase de las surgidas en el decurso evolutivo, nos sobran en cualquier caso motivos para avenirnos al *dictum* de Kandel y conceder que habrá de ser de cuño biológico su explicación –no dejemos de incidir en que «biología» y «explicación» han de leerse aquí en un sentido amplio y plural, respectivamente–. Pero, ¿en qué consiste "explicar" en biología? Desde luego, en nada parecido a lo que cabría pretender que les cupiera hacer o contribuir a hacer a las soluciones separatista e inseparatista. Para avanzar hacia la explicación de las diversas clases de metabolismo hubo que atender a los procesos que se hallan a la base de las mismas y no bastó, sobra decirlo, con analizar, al modo de separatistas e inseparatistas, sus rasgos comunes o con especular acerca de las condiciones formales que algo ha de cumplir para contar como metabólico: el análisis de *nuestra* noción de metabolismo y los principios necesarios para la explicación de las cascadas bioquímicas en que las diversas clases de metabolismo consisten son dos cosas entre las cuales difícilmente podría mediar un trecho mayor. Nada se explica en biología mediante el solo análisis a priori y difícilmente puede el mismo contribuir a elevar el edificio de una explicación biológica. El análisis de *nuestra* noción de representación en que se basan las soluciones separatista e inseparatista no pasa de constituir una especulativa descripción mínima de un conjunto convencional difuso: el integrado por todo aquello que denominamos representacional (téngase presente que cada autor denomina así aquello que tiene a bien). Pero el análisis especulativo de nuestro concepto de digestión pudo contribuir en escasa medida a la explicación de la digestión. Pudo servir, cuando más, para abstraer y listar rasgos compartidos por las diferentes clases de digestión y estable-

cer así taxonomías y definiciones de un valor, por decir lo menos, relativo, o para improvisar condiciones formales de pertenencia a clase y categorías de dudosa potencia heurística y exigua fecundidad explicativa.

Por otra parte, parece haber calado hondo en la bibliografía contemporánea la idea de que la solución al problema de la conciencia, es decir, la explicación científica de la misma, es algo que puede lograrse mediante un perspicuo puñado de proposiciones, idea que no podría resultar más desorientadora: el día en que una explicación de la experiencia consciente se halle disponible no podremos consultar un artículo científico que la contenga, porque la cantidad de datos, niveles explicativos, disciplinas y marcos teóricos implicados en la explicación de cualquier fenómeno biológico es algo que no entra en un artículo –y la experiencia consciente es un fenómeno biológico que no parece exagerado presentar como uno de los más plurales, complejos e interrelacionados, por cuanto se encuentra vinculado con la práctica totalidad de los fenómenos afectivos, cognitivos y conductuales: percepción, atención, memoria, emoción, motivación, funciones ejecutivas, razonamiento, toma de decisiones, conducta social, etc.–. Así, por ejemplo, los procesos a los que hay que atender para explicar la digestión de los rumiantes, o la de los hongos, no son los mismos que aquéllos a los que hay que atender para explicar la digestión de los omnívoros. No existe, entonces, algo así como *la* explicación de la digestión, en primer lugar porque existen diversas formas de la misma en diferentes ramas del árbol filogenético y, en segundo lugar, porque en cada una de esas ramas, al fenómeno que denominamos digestión subyace una plétora de procesos bioquímicos, celulares, fisiológicos, etológicos y hasta ecológicos cada uno de los cuales se explica dentro de diferentes marcos teóricos cuya articulación requiere, desde luego, algo más que un puñado de proposiciones unilateralmente dispuestas para ser llevadas de la urdidera de una sola disciplina al telar de un único paradigma. Al igual que en el caso de la digestión, es más que probable que no quepa esperar *una* explicación de la conciencia, esto es, que quepa antes bien esperar, en primer lugar, que existan diversas clases de conciencia en diversas ramas del árbol filogenético (Denton, 2005: 96 del original, 153 de la traducción) y, en segundo lugar, que sus explicaciones requieran de atenciones diferenciales a elementos diferenciales desde distintos marcos teóricos e involucren así una importante cantidad de niveles explicativos. ¿Qué ha venido haciéndonos a obviar esta plausible posibilidad? El mito de la unicidad, un mito según la cual la conciencia es un fenómeno singular, compacto y homogéneo que habrá de ser explicado mediante el singular, compacto y homogéneo chim-pun largamente anunciado aunque

ya escasamente esperado. Este mito simplista e infundado, endémico en los *Consciousness Studies*, asume que existe *una* explicación de la conciencia, que la misma puede quedar recogida en una serie discreta de proposiciones y que provendrá de investigaciones realizadas dentro de *una* disciplina y ceñidas a *un* marco teórico y un nivel explicativo, momento este último en que el mito de la unicidad desemboca en el campo de batalla de la guerra interparadigmática mostrando el sentido en que ambas querencias teóricas convergen en una concepción simplista de las ciencias y sus objetos. Sin ese hipertrofiado órgano de la imaginación gracias al cual parecen venir muchos anticipando una teoría unificada de lo mental, una teoría cognitiva del todo –habitual aunque ya no exclusivamente galibada de acuerdo con los principios del reduccionismo representacional–, la aludida concepción simplista se vería condenada al ostracismo. Según dicha concepción, el universo de lo mental será finalmente recogido en una teoría omniexplicativa. Pero se da el caso de que tampoco el universo de la física, del que vienen encargándose las disciplinas científicas más maduras, acaba de amoldarse a una sola teoría, sino que es actualmente descrito y explicado por dos teorías cuánticas de campos a nivel de la física de partículas (la teoría electrodébil y la cromodinámica cuántica) y por la teoría general de la relatividad, la métrica de Friedman-Lemaître-Robertson-Walker, la física nuclear, la historia térmica del universo y la nucleosíntesis primordial a nivel cosmológico. Estos marcos teóricos conforman, por otra parte, nada más que los esqueletos de los modelos estándar de la física de partículas y la cosmología. A ellos se suman, más allá de las nubes de hipótesis tentativas que los rodean, una cantidad de leyes, principios, parámetros libres y modelos matemáticos de los que hay que echar mano de forma diferencial en función de aquello que pretendamos explicar.

La respuesta a la pregunta acerca de aquello en lo que consiste una explicación biológica depende, por otra parte, de lo que entendamos por explicación científica. No es éste el lugar para desarrollar una disquisición mínimamente escrupulosa al respecto. Sin embargo, unas superficiales pinceladas se hacen en este punto ineludibles. Desde que Carl Hempel y Paul Oppenheim definieran a finales de los cuarenta la explicación científica como un argumento lógico en el que el *explanandum* (el enunciado que expresa el hecho a explicar) es una consecuencia lógica del *explanans* (el conjunto de enunciados que constituyen la explicación, que habría de incluir, desde esta perspectiva nomológica deductiva, un enunciado que represente una ley científica o, al menos, una generalización legaliforme), la noción de explicación científica ha sido ampliada y ma-

tizada desde perspectivas que exceden el marco positivista lógico dentro del cual fuera inicialmente planteada como "a valid conclusion with which we can explain a fact by beginning with constraints and initial conditions and adding natural laws" (Walter, 1999: 98). Estas ampliaciones y rehechuras no han conducido a una concepción unitaria de la explicación científica, lo cual es particularmente cierto dentro del área de la biología, en la que existe un general consenso acerca de la pluralidad de clases específicas de explicación. Así, por ejemplo, en biología evolutiva y ecología, la forma de la explicación depende a menudo de generalizaciones cuantitativas con base en diferentes modelos matemáticos, que suelen representar en estas disciplinas la dinámica de diferentes sistemas biológicos (modelos de cambio en frecuencias génicas o en el tamaño, estructura y distribución de poblaciones). En la genética de poblaciones, la genética cuantitativa y la genética molecular evolutiva encontramos la base matemática fundamental de la biología evolutiva, pero ello no debe hacernos pensar que exista un único modelo matemático general para la biología evolutiva, pues en realidad lo que tenemos delante en este punto es un conjunto de diferentes modelos matemáticos superpuestos. Dentro de este marco plural cabría esperar que el cambio evolutivo se erigiera como núcleo unificador de los señalados modelos, dado que existe un acuerdo generalizado según el cual el mismo sólo se produce en presencia de variaciones hereditarias y de la presión ejercida en una u otra dirección por la selección natural, que determinaría una reproducción diferencial de las aludidas variaciones. Sin embargo, se da el caso de que tales variaciones pueden concebirse como encarnadas en los genes, los rasgos fenotípicos o, incluso, en segmentos concretos de poblaciones de organismo, de forma que cada una de estas encarnaciones pueden ser objeto de selección y verse sometidas a cambios evolutivos y, adicionalmente, la selección puede ocurrir en varios niveles simultáneamente, con lo cual, la mayoría de los modelos tratan de capturar sólo algunos aspectos del proceso evolutivo y se hallan así bajo el influjo de un dominio particular y, por tanto, lejos de la aspiración de elaborar un modelo matemático unificador que pudiera servir como base de un concepto unificado de explicación (Brigandt, 2013). La imposición de una concepción plural de la explicación biológica viene, por otra parte, dada por el carácter estadístico de la teoría de la evolución, que en su forma contemporánea acepta la base estocástica de la deriva genética, fundada en efectos de muestreo aleatorio. La referida imposición de pluralidad tendría en este punto que ver con el hecho de que no cabe esperar explicar fenómenos de esta índole en los mismos términos que fenómenos fisiológicos, que parecerían plegarse antes a una forma mecanicista de explicación. De este

modo, en áreas como la fisiología o la biología molecular, suele aceptarse que la forma apropiada de explicación depende del descubrimiento de fenómenos sometidos a una forma causal mecanicista de funcionamiento que poco o nada tendrían que ver con aquéllos que en ecología o biología evolutiva tratan de explicarse partiendo de las señaladas generalizaciones cuantitativas y estadísticas. No obstante, también las nociones de causalidad y mecanismo se han visto sujetas en el contexto de la explicación biológica a una consideración plural y han sufrido una historia de hechuras y rehechuras desde el planteamiento reduccionista de Salmon, que en su concepción de la explicación causal –en cuya evolución no podemos entrar aquí (vid, Salmon, 1984; 1998)– definió la causalidad como una propiedad objetiva de los procesos que describe la física, hasta la reelaboración de la noción de mecanismo desarrollada por Bechtel (2006; 2008). A pesar de que las explicaciones causales y mecanicistas han sido a menudo tratadas de forma indistinta en términos de concepción causal-mecanicista de la explicación científica (vid., v. g., Salmon, 1998), lo cierto es que las diferencias entre ambas perspectivas se hacen notorias de lieve y han de ser así puestas de relieve por cuanto conducen a interpretaciones muy diversas del significado de la voz «explicación». Así, en el análisis de la noción de mecanismo propuesto por Bechtel aparece la misma caracterizada como una estructura que desempeña –al menos– una función en virtud de sus partes componentes, sus operaciones componentes y su organización (Bechtel & Abrahamsen, 2005; Bechtel, 2006; 2007; 2008: cap. 1). Un mecanismo sería en este sentido un sistema *organizado*, esto es, dotado de un modo orquestado de funcionamiento, tanto en el plano de la extensión temporal como en el de la composición espacial de sus partes, que incluirían no sólo distintas entidades –variables, en función de la actividad del mecanismo, en número, posición y propiedades–, sino habitualmente diferentes clases de entidades y, asimismo, diferentes formas de interacción causal. Partiendo de la reinterpretación de la noción de causalidad ínsita en esta reformulación en clave biológica de la noción de mecanismo, la explicación causal es replanteada como una forma de representar la actividad de los aspectos relevantes de un mecanismo y, así, lo que explican las explicaciones causales mecanicistas sería el trayecto hasta un estado final definido de un mecanismo dado o la clase de comportamiento que se produce regularmente en un tipo concreto de mecanismo, en ambos casos a expensas de una adecuada descomposición y recomposición del sistema que sea el caso en las pertinentes partes o niveles explicativamente relevantes, una descomposición y recomposición que debiera servir para mostrar cómo el comportamiento o las capacidades de un sistema complejo resultan de la

interacción de sus partes sin que ninguna de ellas se comporte como el sistema todo o posea sus capacidades. Esta descomposición y recomposición mecanicista implica romper con la idea de la autonomía de los niveles explicativos: el objetivo es conseguir una comprensión de la organización integrada de los diferentes niveles en que cabría descomponer el mecanismo, una comprensión del modo en que la actividad orquestada de los fenómenos acaecidos en los diferentes niveles componentes dan lugar a los fenómenos observables al nivel del mecanismo entendido como un todo. La explicación mecanicista es, pues, una forma integrada de explicación, una explicación internivel que implica un descenso del nivel global del mecanismo al de sus partes componentes integrado con su correlativo ascenso de éstas a aquél (Bechtel, 2009). Esta ruptura mecanicista con la idea de la autonomía de los niveles explicativos viene a oponerse frontalmente a la ortodoxia funcionalista en las ciencias cognitivas: mientras para ésta los niveles explicativos se encuentran aislados y, particularmente, mientras para ella el así llamado nivel de implementación podía concebirse como irrelevante por lo que a la explicación de los fenómenos mentales se refiere, la perspectiva mecanicista busca integrar los diferentes niveles de explicación mostrando el modo en que las propiedades del mecanismo se hallan implementadas en sus componentes.

Todo rasgo biológico puede y debe ser, en cualquier caso, explicado desde dos perspectivas. En primer lugar, debe explicarse por qué existe el rasgo en cuestión, cosa que no cabe hacer sino partiendo de la filogénesis.[6] Esta clase de explicación biológica tiene pues un obvio carácter histórico, un carácter que no encontramos en las explicaciones de los fenómenos de los que se encargan la física o la química, dado que los mismos, a diferencia de los seres vivos, no deben sus características, en un sentido explicativamente relevante, a su historia. Explicar, desde este punto de vista histórico, significa desentrañar el modo en que un rasgo surge en la filogénesis y los motivos por los cuales su herencia diferencial provoca su mantenimiento o perpetuación a través de sucesivas generaciones. En segundo lugar, debe explicarse cómo funciona o tiene lugar el fenómeno o rasgo que sea el caso, lo cual implica a menudo –y en virtud de la complejidad del *explanandum*– niveles explicativos que van del molecular, celular, fisiológico y etológico al ecológico. Podemos denominar histórica a la primera perspectiva y funcional a la segunda.

[6] "'Why?' questions deal with the historical and evolutionary factors that account for all aspects of living organisms that exist now or have existed in the past" (Mayr, 1997: 115).

Ambas perspectivas pueden ensayarse sobre fenómenos, procesos y rasgos que cubren toda la gama de la complejidad y la generalidad: desde la embriología del equidna a la neuroetología de la sexualidad de la rata topo lampiña, desde la filogénesis del ojo y las fisiologías de la visión hasta las del altruismo. Por otra parte, podemos alcanzar a desarrollar con relativo éxito una de ellas sin obtener resultados análogos en la otra. Así, recurriendo a un ejemplo de alta generalidad y alta complejidad, la filogénesis de la sexualidad es a día de hoy la arena de debates irresueltos y, de este modo, "la existencia del sexo supone un enigma para los biólogos evolutivos" (Carmona, 2011: 14) mientras, por su parte, la explicación funcional de la sexualidad, esto es, nuestra comprensión de los mecanismos fisiológicos que la sustentan, se encuentra a día de hoy en menor medida a la deriva, es decir, en menor medida a merced del desacuerdo y las conjeturas de difícil operacionalización y, por tanto, confutación. En el caso de la conciencia, las conjeturas y el desacuerdo son prácticamente lo único que encontramos, tanto por lo que a la primera como por lo que a la segunda perspectiva explicativa respecta.

En cuanto a la perspectiva explicativa histórica, el modo en que surge cualquier rasgo en la filogénesis es azaroso: el origen está siempre en una mutación –o, de acuerdo con Lynn Margulis, en la "adquisición de genomas", principalmente por vía simbiogenética (vid. Margulis & Sagan, 2002)–. Por lo que se refiere a la perpetuación de lo nacido de ese azaroso origen, la idea de función da la clave: cada rasgo obedece, obedeció u obedece actualmente aunque no necesariamente de forma idéntica en su origen –el matiz disyuntivo temporal es importante, habida cuenta no sólo de la extendida presencia de órganos vestigiales, sino también de las críticas a la ortodoxia adapticionista desde la distinción entre rasgos adaptativos procedentes de rasgos en su origen igualmente adaptativos y rasgos adaptativos de origen en este sentido diverso– a una capacidad ampliada para dejar descendencia. En otras palabras, la inmensa mayoría de los rasgos biológicos tienen su justificación filogenética en la función que cumplen o cumplieron en su momento. Así, cabe dividir la explicación histórica en dos subclases: la que atiende al origen, poco menos que indescifrable, y la que atiende a la perpetuación. Dado el carácter de la primera, marcado no sólo por su concreción histórica sino también por el papel que juega el azar, por ejemplo, en la producción de las mutaciones, es la segunda la única que parece sernos hacedera. La misma ha de tratar de fundamentar filogenéticamente la respuesta a la pregunta acerca de la función que cumple o vino a cumplir un determinado rasgo. Por su parte, la explicación funcional trata de responder a la pregunta acerca de cómo funciona, es decir, la pregunta acerca del mecanismo en virtud del

cual puede el rasgo que sea el caso desempeñar su función. Se trata, respectivamente, de las siguientes preguntas: ¿Qué función tiene? ¿Cómo la desempeña?

Hemos apuntado ya que actualmente disponemos de una explicación funcional razonablemente bien resuelta junto a una explicación histórica irresuelta en el caso de la sexualidad. Esto podría hacernos pensar que se trata de dos ámbitos explicativos totalmente independientes. Ciertamente, una buena explicación histórica puede ofrecer indicaciones acerca del camino a seguir en la elaboración de determinadas explicaciones funcionales, pero desentrañar el mecanismo en virtud del cual un rasgo dado alcanza a cumplir su función es algo que no parece requerir en principio de ningún conocimiento acerca de su historia filogenética. Así, por ejemplo, parece que podemos abordar la fisiología de, digamos, el estro sin preocuparnos por su historia evolutiva. Sin embargo, conforme avanzamos de la célula al órgano y de éste al sistema, las nociones históricas adquieren un peso creciente dentro del marco de nuestro discurso acerca de los rasgos biológicos de que se trate. A nivel molecular, la explicación de los fenómenos físico-químicos acaecidos en la deshidratación del ácido cítrico no requiere de ninguna mención tácita o explicita a la historia filogenética. A nivel celular, la explicación de los mismos fenómenos, aun cuando no requiera del recurso explícito a nociones históricas, las lleva ya implícitas: sin ir más lejos, el ciclo de Krebs con el que arranca la señalada reacción tiene lugar, en eucariotas, en la matriz mitocondrial, de cuya incardinación en la célula eucariota darían cuenta, de hallarse la teoría endosimbiótica en lo cierto, fenómenos históricos que tuvieron lugar hace 1.500 millones de años. En verdad, da la sensación de que podrían recorrerse ambas rutas explicativas de forma independiente, y la de que ambas guardan un circunspecto silencio mutuo. No obstante, imaginémonos trazando diagramas de conexión que impliquen proyecciones axónicas del laríngeo recurrente a fin de elaborar un modelo de la fisiología de la fonación de la jirafa. Sólo la atención a la filogénesis, sólo la mirada histórica, pues, da cuenta del sentido del sinsentido –desde el punto de vista del "diseño"– a que obedece la dilación, que habría de incardinar explicativamente dicho modelo, provocada por el rodeo de casi cinco metros que este nervio da en esta especie (vid. Dawkins, 2009: 360 y ss. del original, 322 y ss. de la traducción). Más aún, a diferencia de los fenómenos estudiados por las ciencias físicas, y dado que los seres vivos tienen las características que tienen por cuanto ancestros suyos con esas mismas características fueron favorecidos frente a variantes sin ellas, en biología, pero no en física, la respuesta a la pregunta acerca del cómo, a la pregunta acerca del modo en que un determinado fenómeno tiene lugar, implica en buena medida

la respuesta a la pregunta atinente al por qué, a los motivos por los cuales el mismo se vio favorecido en el proceso selectivo. La noción de función vincula ambas preguntas: cómo y por qué. Para explicar cómo funciona un determinado rasgo debemos comprender su función, y esto sólo es posible si nos encontramos en disposición de responder a la pregunta acerca de los motivos por los cuales el mismo se vio favorecido en el proceso selectivo, pues sólo a la luz de la filogénesis cobra sentido la noción de función. Por su parte, la de mecanismo sirve asimismo para hacer explícito este nexo entre cómo y por qué. Los seres vivos, en tanto que sistemas alejados del equilibrio termodinámico activamente implicados en el mantenimiento de su organización, pueden ser concebidos como mecanismos biológicos, esto es, como sistemas complejos cuyo comportamiento se debe a la interacción organizada de sus partes, y dado que tal comportamiento es así resultado de dicha organización, un ser vivo, en tanto que mecanismo, no es sino una solución al problema de organizar sus partes y las interacciones entre las mismas de tal modo que alcance a producir la forma apropiada de comportamiento, es decir, la forma de comportamiento que le permita seguir manteniéndose alejado del equilibrio termodinámico. Tal y como Tinbergen (1963) señalara, causando desde que lo hiciera escaso disenso, nuestra comprensión del modo en que algo semejante es posible requiere de una armonizada combinación de, al menos, explicaciones históricas y funcionales. Las explicaciones funcionales, por otra parte, deben ser comprendidas dentro de determinado orden relacional y jerárquico: se trata de explicar el modo en que diferentes subsistemas interactúan de cara a alcanzar el señalado fin de mantener al organismo alejado del equilibrio termodinámico. La jerarquía partiría del modo en que los sistemas de órganos cumplen a tal fin su cometido, descendería hacia los órganos concretos y el modo en que cada uno de ellos desempeña su función permitiendo así que el nivel anterior haga lo propio, y así sucesivamente, descendiendo peldaños funcionales hasta alcanzar un nivel en el cual la noción de organización deja de jugar un papel relevante y nos encontramos ante fenómenos netamente fisicoquímicos. La comprensión de cada uno de los fenómenos circunscritos en cada nivel funcional, y, particularmente, la de la coordinada trabazón de éstos, necesita, en cualquier caso, de la perspectiva histórica, la única desde la cual, como apuntábamos, cobra sentido la noción de función, del mismo modo en que, tal y como planteara Dobzhansky (1973) en su crítica teísta al creacionismo, sólo a la luz de la evolución cobran sentido los fenómenos biológicos. No nos cabrá, pues, elaborar una explicación solvente de la conciencia que se muestre incapaz de conjugar la perspectiva histórica y la funcional.

¿Y qué es lo que debe explicar una explicación biológica de la conciencia? Desde luego, son muchos los niveles que cabe esperar de una explicación de la conciencia, dado que, por un lado, y tal y como vimos ya en el segundo capítulo, no existe una sola forma de conciencia y, por otro, eso que denominamos experiencia consciente es algo relacionado con la práctica totalidad de los fenómenos que denominamos "mentales". No obstante, entendemos que, en el primer nivel, en el sentido fundamental, el *explanandum* de una explicación de la conciencia "is the capacity of any organism to sense or to feel something –anything" (Macphail, 2000: 253), de modo que el primer nivel del *explanans* habrá de atender al valor adaptativo de tal capacidad de sentir o experimentar. Con esto, un obvio prerrequisito para la elaboración de una explicación de la experiencia consciente acorde con los lineamientos trazados pasaría por un serio intento de superar las asunciones implícitas en el uso del vocabulario endémico en la disciplina. Hemos dedicado ya suficiente espacio a menoscabar esa intuición no tematizada según la cual cada una de las partes de la consuetudinaria contraposición entre lo fenoménico y lo representacional referirían unívocamente a fenómenos ontológicamente homogéneos, discretos y aproblemáticos. De forma tácita o explícita, esta intuición viene acompañada en la bibliografía contemporánea de otra según la cual si nos encontramos ante determinado aspecto fenoménico, entonces no puede el mismo ser al tiempo intencional, representacional o funcional y, a la inversa, si nos hallamos frente a determinado aspecto de la economía funcional de lo mental, entonces no puede ser el mismo al tiempo fenoménico. Nuestra apuesta, perieca de esta tendencia ortodoxa que Güven Güzeldere ha denominado segregacionismo (Güzeldere, 1997: 11), es la de Darwin, esto es, la apuesta según la cual podemos contar con los dedos de una mano todas las situaciones más improbables que ésa en que la conciencia habría evolucionado sin cumplir ninguna función (James, 1892: 99). Se trataría, en otras palabras, de la apuesta según la cual el aspecto fenoménico de la conciencia nunca podrá ser científicamente explicado si lo consideramos de forma aislada respecto de la función biológica que cumple, es decir, respecto de lo que el mismo hace. Asumiendo que carece de sentido concebir un rasgo biológico tan notable como la conciencia como un subproducto epifenoménico sin valor adaptativo –aunque sin descartar que su valor adaptativo pudiera proceder en determinadas líneas filogenéticas de lo que inicialmente fuera un subproducto derivado del entrecruzamiento de otros caracteres adaptativos o, incluso, no adaptativos–, lo que toca preguntar es, justamente, cuál hemos de entender que es la función biológica que desempeña. La respuesta a tal pregunta habrá de referir al modo en que la experiencia

consciente contribuye al control de la conducta, dado que "la conciencia sólo puede tener valor adaptativo y función biológica en virtud de su capacidad para influir en el comportamiento" (Earl, 2014: 2τ). Pero, ¿de qué modo puede influir en la conducta la experiencia consciente? Aunque, sin duda, formular esta pregunta en singular simplifique excesivamente la cuestión, responder a la misma pasa por la elucidación de los dos siguientes extremos: por qué evolucionó la conciencia y qué es lo que hace. Comencemos por el primero.

### 2.1._La función de la experiencia I. Por qué evolucionó la conciencia

Si la conciencia es un rasgo biológico, o una plural colección de ellos, o incluso una abigarrada clase de complejos sistemas de rasgos biológicos, entonces debe haber una explicación evolutiva de su presencia en el mundo biológico. Anotemos de pasada que afirmaciones como ésta no nos comprometen con la ortodoxia adaptacionista. Bien cabe que no todas las características de los seres vivos puedan explicarse como adaptaciones, que no todas surgieran y se mantuvieran sobre la exclusiva base de procesos de adaptación regidos por la selección natural, pero son ciertamente escasos los ejemplos de rasgos mínimamente complejos actualmente concebidos como explicables desde una perspectiva diversa de la adaptacionista y, de este modo, todo parece sugerir que una explicación evolutiva de la presencia en el reino animal de experiencia consciente habrá de presentarla como producto de la selección natural, aunque muy probablemente como un producto plural en sus formas, estructuras, procesos y linajes. En cualquier caso, si la experiencia consciente debe ser concebida como producto de la selección natural, el epifenomenalismo queda excluido: nada sin potencia causal puede intervenir en procesos de selección natural. Este es un punto crucial: descubrir la(s) forma(s) de la potencia causal de la experiencia consciente enhila la determinación de la(s) función(es) que cumple, y esto, a su vez, la elucidación de los motivos por los cuales la selección natural operó sobre ella. En las antípodas del antifenomenalismo y el epifenomenalismo, la premisa fundamental de una cabal biología de la conciencia podría enunciarse como sigue: la experiencia consciente irrumpe en la historia natural por cuanto habilita el ejercicio de formas de conducta irrealizables –o inoptimizables– en su ausencia y que significaron una importante diferencia de aptitud biológica.

Antes de tomar en consideración los posibles motivos por los cuales la experiencia consciente fue seleccionada conviene ofrecer réplica a quienes, como el epifenomenalis-

ta, suponen que no los hay, esto es, a quienes han venido tratando de injerir visos de verosimilitud en la idea de que la misma no cumple ninguna función biológica. Cabe esgrimir dos clases de argumento en defensa de una tesis como ésta. En primer lugar, desde cierta suerte de paralelismo epifenomenalista, puede argüirse que la conciencia y la conducta se encuentran desconectadas y por tanto la conciencia no puede tener valor biológico alguno. A pesar de la manifiesta implausibilidad biológica de semejante aserto, la fuerza de posturas metafísicas como la que trata de sustentar reside en su carácter indemostrable: nadie puede demostrar su falsedad, pero es precisamente esta propiedad la que determina que esta clase de enunciados caigan fuera del ámbito de cualquier investigación que se proponga avanzar hacia una comprensión prudente y fidedigna de la naturaleza de su objeto. En segundo lugar, puede argüirse que no todos los rasgos biológicos existentes en el presente son, estrictamente, el resultado de la selección natural. Esto sucede, por ejemplo, en la pleiotropía –el efecto fenotípico de un gen sobre varias características–, o cuando el rasgo en cuestión es una consecuencia lateral del utillaje físico del rasgo verdaderamente adaptativo. Así, por ejemplo, puede decirse que el color de la sangre es un subproducto de la estructura de la hemoglobina y que mientras la capacidad de fijar oxígeno es lo realmente adaptativo el "rasgo" biológico «color rojo de la sangre» sería una mera consecuencia tangencial sin valor adaptativo. Pero lo cierto es que esta clase de tangencialidad evolutiva afecta a rasgos biológicos simples o accesorios al punto que resulta conflictivo denominarlos rasgos y, por tanto, sería extremadamente improbable que pudiera la misma dar forma a rasgos biológicos tan complejos como cabe entender que es la conciencia. Hay que admitir, no obstante, que a día de hoy carecemos de evidencia empírica o argumentación racional capaz de bloquear enteramente este tipo de desafío epifenomenalista. Además, demostrar que un determinado rasgo ha evolucionado dado su carácter adaptativo implica, entre otras cosas, probar que el mismo es heredable y que su herencia ha sido a lo largo de sucesivas generaciones diferencial, esto es, que sobre la misma ha actuado la presión selectiva, lo cual sólo es posible mediante evidencia fósil y/o experimental.[7] Tanto en el caso del primer requerimiento con en el del segundo nos movemos a día de hoy en una oscuridad que apenas

[7] De hecho, según Brandon (1990: 165-174), tal demostración requeriría: a) evidencia de que la selección ha tenido lugar, b) una explicación ecológica de la adaptabilidad relativa, c) evidencia de que los rasgos en cuestión son hereditarios, d) información sobre estructuras poblacionales y, por último, e) información filogenética sobre la polaridad rasgo. Sin embargo, son minoría las aproximaciones adaptacionistas que se ajustan a este ideal explicativo. Así, en el caso de la conciencia, una aproximación adaptacionista a la misma no parece peligrar excesivamente al dar por supuesta su polaridad, asumiendo que los organismos conscientes provienen de antepasados no dotados de conciencia y no a la inversa, ni al guardar un circunspecto silencio sobre estructuras poblacionales en entornos adaptativos.

logramos figurarnos desleíble. Pero no es esta oscuridad el verdadero problema cuando de lo que se trata es de atajar el desafío epifenomenalista, dado que el mismo no se funda en el intento de abrir vías a nuestra comprensión de este o aquel fenómeno biológico, sino en el de obliterarlas y salvaguardar una determinada concepción metafísica del psiquismo. No hay evidencia ni a favor ni en contra del teísmo, sino sólo un hatajo de argumentos abiertos a sucesivas reinterpretaciones, pero nunca a su definitiva confutación. Lo mismo sucede con el epifenomenalismo. Todo rasgo biológico mínimamente complejo es producto de la selección natural y cumple una función. Habida cuenta del tipo de actividad biológica que, según todos los datos experimentales y modelos teóricos disponibles, sustenta la experiencia consciente –muy probablemente una de las formas más complejas de actividad biológica–, empeñarse en que la misma constituya la excepción a esta regla sirve para engrosar la ya abultada lista de excentricidades metafísicas occidentales, pero en ningún caso para avanzar de forma justificable hacia una comprensión científica de la naturaleza de la experiencia consciente. Sin embargo, podemos apelar cuanto queramos a la sana sindéresis, que si el epifenomenalista se empeña en sostener que nada puede refutar su tesis, habremos de convenir, cosa que no necesariamente hablará en favor de la misma (vid, v. g., Popper 1934/1959). No parece haber argumento o evidencia capaz de morigerar ciertas suertes de azacanamiento de raigambre metafísica.

Decíamos poco más arriba que nuestra apuesta es la de Darwin, pero citábamos a James. Quizá hubiéramos debido decir, sin más, que la nuestra es la apuesta de James, pues entendemos, con él, que la conciencia es no sólo un fenómeno biológico originado en el curso de la evolución y que, por tanto, cumple –al menos– una función, sino que, además, se encuentra distribuida a lo largo y ancho del reino animal en diversas formas y grados, y esto fue él el primero en defenderlo mediante argumentos, aunque otros darwinistas hubieran avanzado la idea antes de su sistematización. Así, Thomas Huxley –a cuyo epifenomenalismo se opondría frontalmente el programa funcionalista jamesiano– había afirmado tres lustros antes de la publicación de los *Principles* de James:

> The doctrine of continuity is too well established for it to be permissible to me to suppose that any complex natural phenomenon comes into existence suddenly, and without being preceded by simpler modifications; and very strong arguments would be needed to prove that such complex phenomena as those of consciousness, first make their appearance in man (Huxley, 1874: 236).

James comienza a elaborar su planteamiento acerca de la evolución de la conciencia cuatro años después de que Huxley publicara el texto que acabamos de citar. Dicha elaboración partiría de los bocetos preparados con motivo de una serie de conferencias que impartiera en el Instituto Lowell, unos bocetos que acabarían integrados en los *Principles*. Ya en su primera formulación, James pone en relación su idea de una evolución temprana y multiforme de la conciencia con la filogénesis del sistema nervioso y la complejidad del repertorio conductual de los organismos. Las conclusiones de James coinciden en buena medida con las actuales: existen múltiples formas y grados de conciencia distribuidos por todo lo ancho y largo del reino animal, y todos ellos se encuentran inextricablemente ligados a la filogénesis del sistema nervioso y los repertorios conductuales. No obstante, existe hoy un acuerdo generalizado según el cual mamíferos y aves son fenoménicamente conscientes, pero las dudas y el disenso reinan por lo que al resto de los vertebrados y los grandes cefalópodos respecta. A pesar de esto, las conclusiones que Peter Århem, Hans Liljenström e Ingemar Lindahl extraen de un *workshop* sobre la evolución de la conciencia celebrado en el verano de 2001 en el que participaran, entre otros, Bernard Baars, Jean-Pierre Changeux y Antti Revonsuo, podrían perfectamente haber sido enunciadas por James: "it might become important to distinguish between different degrees and levels of consciousness, related to the complexity of the nervous system of the respective species" (Århem, Liljenström & Lindahl, 2002). [8] Con todo, nosotros solamente podemos asomarnos a esas formas y grados a través de la conducta y la fisiología. El único medio de que a día de hoy disponemos para escrutar la filogénesis de esos diversos grados y acercarnos así al flanco explicativo histórico es, pues, la psicobiología comparada, dado que el registro fósil no prodiga vestigios de conducta ni de órganos blandos.

Es más que más probable que la filogénesis de la conciencia haya venido transcurriendo a lo largo de diversos ríos filéticos y cladísticos y pueda encontrarse así distribuida a lo largo de los mismos en formas homólogas y análogas.[9] Tratar de decidir si la conciencia es un rasgo homólogo o más bien análogo sería plantear la cuestión en términos excesivamente simplistas: la proliferación de ramas en el árbol filogenético y la

[8] Pero, no nos engañemos, en ese *workshop*, al igual que en cualquier otro que pudiera celebrarse hoy con la evolución de la conciencia como tema, la frontera inferior de la misma osciló más allá de la franja entre agnatos y mamíferos.

[9] En sentido biológico, dos estructuras orgánicas de diferentes especies son homólogas si tienen un mismo origen evolutivo –sería el caso de nuestros brazos y las aletas del delfín– y análogas si, ofreciendo respuesta al mismo problema biológico, tienen orígenes diversos –sería el caso de las alas de los insectos, las aves y los murciélagos.

complejidad del fenómeno –al que parece acertado referirse antes bien como a una malla de fenómenos– invitan más bien a pensar en un marco en el que ambas categorías se encuentran solapadas de diversos modos en distintos puntos de diferentes ramas. Al igual que evoluciona específicamente la fisiología en relación con nichos ecológicos particulares, asimismo cabe esperar que suceda con las características mentales, pues nada sugiere que los fenómenos biológicos que denominamos "mentales" y aquéllos que denominamos "fisiológicos" hayan de encontrarse disparejamente sujetos a los principios que describe la teoría de la evolución y quepa así dudar "que la mente sea producto, asimismo, de la evolución" (Rubia, 2000: 50).[10] Por tanto, una *primera hipótesis* plausible[11] sería la siguiente: cabe esperar que la conciencia aparezca, allí donde haya podido contribuir a aumentar la capacidad de los organismos para adaptarse a sus medios, en diferentes ramas del árbol evolutivo en distintos grados y variedades y en función de las soluciones ofrecidas por la selección natural para las particulares demandas que los distintos nichos han venido imponiendo a sus especies. Una explicación de la conciencia humana que no tuviera en cuenta su linaje sería sólo media explicación. Una reconstrucción acabada de la historia de este linaje es, por otra parte, una tarea inabordable. Disponemos, como apuntábamos, sólo de los barruntos que de la historia filogenética puede ofrecernos una atenta consideración psicobiológica de las especies actuales. Sin embargo, lo cierto es que la reconstrucción de la historia filogenética de prácticamente cualquier rasgo biológico se enfrenta a dificultades idénticas (ausencia de evidencia fósil relevante, dificultades en la realización e interpretación de estudios experimentales comparativos, etc.), pero "se trata de las complicaciones con las que los biólogos e historiadores naturales se topan a diario y no reflejan ninguna clase de problemas especiales propios del estudio de la conciencia" (Polger, 2007: 72τ).

El principal obstáculo que parece verse obligada a vencer la atenta consideración psicobiológica de las especies actuales destinada a vertebrar la perspectiva explicativa histórica es, según una extendida opinión, que el intento de sentar las bases para una reconstrucción de la evolución de la conciencia no puede disponer de los medios al alcance de la fisiología evolutiva o la anatomía comparada, cuyas fuentes de datos son directamente observables, sino que dicha reconstrucción debe ensayarse, indirectamente, desde la inferencia a partir del comportamiento y la fisiología de las especies animales

[10] Parafraseando a uno de los padres de la nueva síntesis cabría hablar en este punto de "psicofilogénesis" (Rensch, 1972).
[11] Usamos el adjetivo en el sentido que en este contexto le da Bunge (1967/2004: 223).

contemporáneas, única fuente directa de datos en este punto. Preguntémonos, sin embargo, si constituye un problema tan enjundioso como ha venido pretendiéndose que debamos acudir a la guía de la conducta y la fisiología para alcanzar de forma científica lo mental. "La conciencia es públicamente inobservable, y la observación pública el garante de la cientifidad". Si esta clase de verdad a medias, supuestamente enfrentada al estatus científico del estudio de lo mental, fue esgrimida en una vena positivista ramplona contra la posibilidad misma de elaborar una ciencia de lo mental cuando ésta alboreaba, a nadie debe extrañar que retorne hoy en forma de desaire, de desestimación a priori de una labor apenas emprendida. Si durante la hegemonía conductista resultaba para la inmensa mayoría indudable el estatus acientífico de la experiencia consciente, que jamás puede hacerse pública en sí misma sino sólo, como el Apolo de Heráclito (vid. fragmento 14, en Bernabé Pajares, 2001: 131), a través de las señales que nos prodiga, este rebrote de escepticismo acerca de las posibilidades con las que contaría el proyecto científico apenas incoado de explicar la experiencia consciente sugiere que en una situación mucho peor habrá de encontrarse el intento de reconstruir su filogénesis. En esta línea, ha venido argumentándose, nuevamente desde un trasnochado positivismo, que un requisito para la evaluación empírica de una teoría científica es que la misma contenga hipótesis referidas a fenómenos pública y directamente observables, cosa impracticable en el caso de la conciencia, particularmente la de animales no humanos, que ni siquiera pueden informar verbalmente de sus experiencias conscientes. Sin embargo, las hipótesis científicas no siempre se refieren a fenómenos pública y directamente observables, sino que, en muchas ocasiones, permiten derivar consecuencias observables (el descubrimiento del bosón de Higgs a partir de mediciones de la energía y direccionalidad de fotones o la propia noción de gen pueden servir como ejemplos, pero en realidad cualquier investigación científica que haga uso de cualquier clase de constructo teórico, esto es, la práctica totalidad, podría hacerlo igualmente). Así las cosas, nada parece bloquear a priori la tarea de elaborar una explicación científica de la conciencia y, en lo que al caso que ahora nos ocupa se refiere, es evidente que a fin de trazar la filogénesis de la conciencia debemos partir del establecimiento de las características de la fisiología y el comportamiento de especies animales actuales que cabe entender como expresión de la misma, una expresión cuya taxonomía nos será luego dable incardinar en modelos de su evolución en virtud de los ya elaborados para la filogénesis de especies y fisiologías. En este marco, la labor del científico será, como lo fue siempre, y tal y como el psicólogo norteamericano Ernest Ropiequet Hilgard señalara en su

consideración de las implicaciones que el retorno cognitivista a la conciencia tendría para la psicología, la de utilizar las mejores técnicas disponibles para la verificación de los datos y para la validación de las inferencias realizables a partir de los mismos (Hilgard, 1980: 15τ). Pero se trata de una labor que no tenemos motivos para suponer realizable de forma unilateral, desde dentro de los márgenes de un único paradigma y una sola disciplina, una labor para la cual "el cuidadoso análisis de conceptos y argumentaciones por parte de la filosofía cognitiva será esencial, pero no suficiente; como asimismo será el análisis neurobiológico de los fundamentos fisiológicos y estructurales muy importante, pero tampoco suficiente; del mismo modo que resultará el meticuloso análisis etológico de un amplio espectro de comportamientos animales esencial, aunque, nuevamente, no suficiente" (Hendrichs, 1999: 99τ).

Disponemos de dos medios para el estudio del comportamiento animal que la tarea de trazar la filogénesis de la conciencia exige. Por una parte, en la línea de la tradición griffiniana en etología cognitiva,[12] cabe observar el comportamiento en el hábitat del animal sin intervenir en el mismo ni manipularlo. Por otra, en la línea de la tradición experimental en psicología comparada, cabe estudiarlo tratando de ejercer control sobre las variables consideradas de interés y manipulándolas de forma sistemática para hallar relaciones entre ellas y probar hipótesis acerca de las mismas. La comparación de la metodología observacional con la experimental siempre conduce a este resultado: sin control experimental podemos obtener registros de gran interés y potencial heurístico, pero son escasas las inferencias que de forma segura nos cabe realizar partiendo de los mismos. De cara a trazar la filogénesis de la conciencia, la compilación de un cuerpo de datos obtenidos en el medio natural de los animales estudiados será sin duda de gran ayuda, pero el vínculo buscado entre la conducta y la conciencia difícilmente podrá asentarse de forma exclusiva en semejante base: si tratamos de dar cuenta de la conducta observada apelando a la actividad consciente del organismo, y si lo que pretendemos con ello es realizar atribuciones legítimas de mentalidad consciente en base a criterios conductuales y fisiológicos, la única forma de hacerlo de forma científicamente irreprochable será diseñando experimentos en los que el control sistemático de variables específicas nos permita alcanzar un acuerdo razonado acerca de esos mismos criterios. No obstante, no pretendemos defender la pertinencia de una total desatención a la tradición

[12] Disciplina que arranca formalmente con la publicación en 1976 del influyente texto de Donald R. Griffin *The Question of Animal Awareness: Evolutionary Continuity of Mental Experiencice* y que podemos definir como el estudio comparado, evolutivo y ecológico de la inteligencia y la conciencia animal (Allen & Bekoff, 1997: ixτ).

griffiniana en este punto, porque siendo la conciencia un fenómeno biológico tan vasto y hallándose vinculada con casi cada segmento de la actividad cognitiva, emocional, motivacional y conductual de los organismos, la observación en medios naturales será, con toda probabilidad, más propicia para la investigación de determinados procesos implicados en la experiencia consciente que pudieran verse contaminados por la artificialidad propia de las situaciones experimentales, como, por ejemplo, los de comunicación.

Nuestra primera hipótesis acerca de la filogénesis de la conciencia es excesivamente general y exige por tanto ulteriores esfuerzos para cerrar el cerco de la evolución de la experiencia consciente. Una *segunda hipótesis* plausible que podría contribuir a ello sería la siguiente: toda forma compleja de comportamiento implica un determinado grado de conciencia, que cabe suponer atenuado y hasta ausente allí donde nos encontremos con mecanismos conductuales sometidos a regímenes rígidos, tales como los ejemplificados por las formas elementales del aprendizaje (asociativas, como el condicionamiento respondiente, y no asociativas, como la sensibilización y la habituación) o las pautas de acción modal –otrora denominadas pautas de acción fija–. Tratar como a un autómata a un animal capaz de desplegar un repertorio conductual mínimamente rico quizá resulte tan injustificado como atribuirle una mente consciente. No obstante, disponemos para el caso de nuestro vecino y el de su mascota de idéntico fundamento en que asentar nuestras atribuciones de mentalidad consciente: acaso, ninguno. Nadie ha resuelto el así llamado problema de las otras mentes.[13] Con todo, interpretar la conducta de mi vecino o la de su dachshund como las de sendos autómatas inconscientes nos suena a algunos a interpretación injustificada. Esperar que la selección natural fije por completo el repertorio conductual que cabría esperar de un reptil del carbonífero sobre la exclusiva base de las formas elementales de aprendizaje y el instinto a las que aludíamos sería una verdadera excentricidad evolutiva, dado que todo apunta que no son ni muchas ni muy complejas las formas de conducta que sólo en relación a las mismas podrían ser cabalmente explicadas –de hecho, es más que dudoso que pueda serlo incluso la conducta de los insectos (Ortega Escobar, 2014: 59)–. El principal problema de

[13] Muy probablemente, el mejor argumento siga siendo hoy, como lo era hace ya treinta años, de tipo inductivo y, en ningún caso, demostrativo (en lo relativo al modo en que usamos aquí este adjetivo: vid. Cíntora, 2005: 112). En términos aristotélico-kantianos: la solución al problema de las otras mentes puede que se mueva de lo *problemático* a lo *asertórico*, pero siempre lejos de lo *apodíctico*. "The best argument against other- minds skepticism is, probably, that, given the non-uniqueness of one's physical constitution and the general uniformity of nature in the biological sphere as in others, it is in the highest degree improbable that one is unique (...) in being the enjoyer of subjective states" (Strawson, 1985: 20).

esta segunda hipótesis reside en el término «complejo». ¿Qué clase de patrones de conducta consideraremos "complejos"? La respuesta, escueta y contenida en lo antedicho: la clase de los patrones de conducta que excedan la potencia explicativa de las herramientas teóricas diseñadas para dar cuenta de las señaladas formas elementales del aprendizaje y el instinto, esto es, la práctica totalidad de la conducta animal.

El reino animal puede dividirse en dos grandes grupos o taxones: el de los animales bilaterales, dotados de una forma bilateral de simetría, y el de los animales radiados, dotados de una forma radial de simetría y más primitivos que los anteriores. Este segundo grupo incluye los filos cnidaria y ctenophora, el primero, por su parte, unos treinta, entre ellos el nuestro, el de los cordados craniados. La simetría bilateral, ligada a las peculiares formas de motilidad que habilita, aparece en todos los organismos triploblásticos, aunque en algunos casos, como el de los equinodermos, llega a perderse en la fase adulta. La característica distintiva de los organismos triploblásticos, noción usada en contraposición a la de diploblástico, es la presencia en el desarrollo embrionario temprano de tres capas germinativas u hojas embrionarias, el endodermo, el mesodermo y el ectodermo, capa esta última que emerge en primer lugar del epiblasto durante la gastrulación y aparece situada así inicialmente en la parte externa de la gástrula: de ella proceden, curiosamente, la piel y el sistema nervioso –nuestros *non plus ultra* interior y exterior–. Los únicos animales que cabía encontrar en Panthalassa hace 700 millones de años eran diploblásticos. Nuestro linaje surge poco después escindiéndose del del resto de metazoos. Hace unos 550 millones de años nuestro linaje vuelve a escindirse: es el momento en que los cordados seguimos nuestro camino y nos desvinculamos de la estirpe de los urocordados. Entre 8 y 20 millones de años después, durante la explosión cámbrica, encontramos ya entre los celomados, animales en cuya embriogénesis el mesodermo se ahueca formando en el interior del animal la así llamada cavidad general, a los protóstomos, en los que la boca del adulto deriva del blastoporo embrionario, y los deuteróstomos, nuestra línea, en la cual no es la boca del adulto, si no su ano, lo que deriva del blastoporo, siendo aquélla de neoformación. Dentro del linaje de los cordados surge el de los peces hace 530 millones de años, primero los agnatos, peces sin mandíbula –grupo parafilético poco representado actualmente (apenas cien especies de agnatos, como la lamprea, comparten hoy con nosotros el planeta), pero decisivo: son los primeros vertebrados, es decir, los primeros craniados, es decir, los primeros animales en proteger su "tesoro, el cerebro, (…) [en] una caja de caudales, el cráneo" (Mosterín,

2006/2008: 92; 2013b: 124)–, luego los gnatóstomos, el clado de los organismos con mandíbula, nuestro clado (en realidad, una superclase de vertebrados que incluye a la inmensa mayoría de peces, anfibios, reptiles, aves y mamíferos actuales). Éste, el de los peces, es nuestro linaje.[14] Sólo nos escindiríamos del mismo hace 400 millones de años con la aparición de los primeros tetrápodos, los anfibios, de cuya parentela nos desligáramos hace 340 millones de años con la aparición de los reptiles. Pasarían 120 millones de años hasta que esta rama se dividiera, dando lugar a la de los mamíferos, de la cual brotara la de los primates hace al menos 65 millones de años. De ella se bifurca la de los homínidos hace unos 2,5 millones de años. La de *Homo sapiens* lo hace de ésta hace unos 300.000 años.

La filogénesis del sistema nervioso coincide con la del reino animalia: todos los animales tienen sistema nervioso. El sistema nervioso es seleccionado y apuntalado en los primeros compases de su historia evolutiva a causa de la necesidad en que los animales nos encontramos de desplazarnos. Los primeros animales semovientes fueron los cnidarios y, congruentemente, en ellos encontramos las formas más primitivas del sistema nervioso: redes de neuronas sin ninguna clase de agrupación anatómicamente diferenciada y funcionalmente especializada. La integración de cuerpos celulares en ganglios llegaría más tarde. En cualquier caso, si descartamos la hipótesis panpsiquista, concederemos que la conducta propositiva de los seres vivos apareció en la historia filogenética antes que cualquier clase de actividad mental, y también antes de la aparición de las células nerviosas. Las de vida e intención son nociones solapadas y difícilmente desligables, pero las intenciones de los seres vivos no dotados de sistemas nerviosos funcionalmente organizados tienen con toda probabilidad poco o nada que ver con intenciones conscientes: la intención de vivir y la conducta propositiva que propicia se encuentran en estos organismos codificadas en el ADN y en nada se asemejan a la clase de actividad biológica que denominamos "mental". Vestigios de los primeros destellos evolutivos de conducta propositiva pueden observarse hoy en las bacterias, que se relacionan de forma activa y adaptativa con su entorno a pesar de carecer de sistema nervioso. Así, por ejemplo, un ejemplar de *Escherichia coli* inspeccionará su medio usando proteínas de membrana a modo de órganos sensoriales y agitará sus flagelos para eludir sustancias tóxicas o desplazarse desde zonas de ausencia o baja concentración de gluco-

[14] Hecho que, al parecer, desconcierta profundamente a los creacionistas. El origen marino de nuestro linaje ha sido así el protagonista de la denominada "batalla de los peces" en Estados Unidos (vid. Eldredge, 2005: 23 de la traducción).

sa hacia zonas de mayor concentración. La conducta propositiva aparecerá en formas cada vez más sofisticadas, pero para que esto sucediera tendrían que aparecer las células eucarioas e instaurar las agrupaciones solidarias y funcionalmente especializadas que encontramos a lo largo y ancho del reino animalia. En los poríferos, el más antiguo entre los filos animales con los que compartimos el planeta, encontramos ya un tejido neuroepitelial responsivo a estímulos táctiles y químicos y gestor de la conducta de los poros a través de los cuales filtran estos motazoos el agua para extraer de ella los nutrientes que constituyen su dieta. Tenemos ya no sólo conducta propositiva, como en la *Escherichia coli*, sino también mediada por la acción de un proto-sistema nervioso elemental e indiferenciado, pero pocos o ningún motivo para apoyarnos en ella y realizar atribuciones de mentalidad.

Según la reconstrucción contemporánea de la historia filogenética del sistema nervioso, el primer peldaño que la misma asciende es el que lleva de la indiferenciación neuroepitelial porífera a la aludida red cnidaria. En ella encontramos por vez primera neuronas propiamente dichas, unas células que apenas han variado a lo largo de su historia evolutiva. En realidad, lo único que ha variado a lo largo de la misma ha sido la complejidad de la organización del sistema nervioso (Swanson, 2012: 40). Este primitivo sistema nervioso ya no es meramente responsivo: ahora emprende acciones, es el motor que permite al organismo explorar su entorno. En la red cnidaria hallamos, pues, la referida prístina función del sistema nervioso: el control del movimiento. Los beneficiarios de esta red nerviosa difusa, animales como las medusas o las anémonas, presentan junto a lo que puede denominarse ya propiamente un tejido nervioso, fibras musculares, glándulas y células sensoriales. El tejido nervioso de esta clase de animales lo componen células bipolares y multipolares no organizadas sistémicamente ni diferenciadas funcionalmente, al punto que no cabe distinguir entre axones y dendritas en sus prolongaciones y los potenciales de acción pueden atravesarlas en ambas direcciones. La estimulación se traduce con rapidez y rigidez en acción. Las células sensoriales, las nerviosas y los órganos efectores de estos animales se encuentran muy próximos: apenas hay lugar para procesos mediadores y de aprendizaje y, de hecho, a pesar de haberse hallado evidencia de condicionamiento clásico en este filo, no existe consenso acerca de si esta forma básica de aprendizaje juega algún papel en la conducta de estos animales en sus medios naturales (Smith, 2008: 72). ¿Encontramos en este primitivo tejido nervioso el primer destello de experiencia consciente? Nadie puede responder a esta pregunta. No obstante, no resulta sencillo desvincular la respuesta afirmativa de cierta suer-

te de compromiso con una concepción mágica del funcionamiento nervioso según la cual la mente surge de él espontáneamente, como si a la célula nerviosa le fuera inherente, como si no fuera un producto de la organización y coordinación del sistema nervioso con el resto del organismo sino de una misteriosa esencia mental connatural a la economía de la célula nerviosa. Del mismo modo que cuesta creer que la "endocrineidad" haya sido algo inherente a la filogenia de determinada clase de célula, y del mismo modo que parece más razonable entender la especificidad funcional de los órganos que encontramos en el presente como gestada a partir de células ni especializadas ni organizadas en su origen del modo en que actualmente se encuentran especializadas y organizadas, cuesta creer que poner una neurona junto a otras cien, o mil, o cien millones, vaya a originar conciencia sin más, y parece más razonable entender los fenómenos mentales conscientes no como surgidos de la evolución de una clase de célula especial, sino de la de una clase de órgano, de una forma de organización: la del sistema nervioso en coordinación con el resto del organismo. De ahí que resulte verosímil concebir los motivos que pudieran movilizar la atribución de mentalidad consciente a organismos dotados de una organización tan elemental como antes emocionales que racionales, mas motivos, en cualquier clase, enteramente loables, pues de algún modo traslucen un instintivo respeto por los seres vivos, bien que mediado por una cierta suerte de narcisismo cinético que convierte en sujeto moral privilegiado al ser vivo semoviente en detrimento del estacionario.

El siguiente prototipo que la selección natural modela para el sistema nervioso es el denominado sistema hiponeuro o ganglionar, presente de forma definida ya en anélidos hace unos 500 millones de años (los fósiles de *Cloudina* son 50 millones de años anteriores a esta datación, pero los taxónomos se dividen entre los que los clasifican entre los cnidarios y los que los clasifican entre los anélidos), prefigurado en protóstomos anteriores como los platelmintos, nemátodos o moluscos y refinado en una línea filogenética paralela a la de los anélidos, la de los artrópodos. En este nuevo diseño, las células nerviosas individuales dejan de constituir el nivel básico de organización cediendo el relevo a una unidad funcional mayor, el ganglio, que integra somas de células nerviosas destinadas a coordinar conjuntamente diversos aspectos de la conducta del organismo. Estas agrupaciones celulares tienden a ubicarse en mayor número rostralmente, en ganglios cerebroides situados alrededor del esófago y conectados con los cordones nerviosos ventrales mediante conectivos periesofágicos. Esta posición ventral del cordón nervioso sería la que llevara al naturalista francés Étienne Geoffroy Saint-Hilaire

a afirmar en la segunda década del XIX que los vertebrados somos anélidos al revés (Gould, 1985: 38 del original, 32 de la traducción; Maynard Smith, 1998: 11 del original, 25 de la traducción). En cualquier caso, y de forma en cierto sentido análoga a la división entre el sistema nervioso central y el periférico, en el sistema nervioso ganglionar puede distinguirse ya una porción central integrada por el conjunto de ganglios que controla la porción periférica, constituida por los receptores sensoriales y los nervios a través de los cuales los ganglios reciben desde aquéllos aferencias y envían y reciben eferencias y aferencias desde y hacia músculos y glándulas.

La filogenia posterior del sistema nervioso sigue un curso que es dable epitomar mediante la voz «encefalización». No todos los ganglios que encontramos en los metámeros de los anélidos presentan el mismo volumen: los que desempeñan funciones de mayor complejidad y relevancia son mayores, y la selección natural ha dispuesto que los mismos tiendan a situarse, como indicábamos, rostralmente, en la parte anterior del organismo. De este progresivo abultamiento de los ganglios rostrales derivan nuestros encéfalos mamíferos, pero la evolución no ha dado origen a nada similar a ellos en invertebrados –excepción hecha de los sistemas nerviosos centrales de los grandes cefalópodos–, sino sólo variaciones sobre los diseños hasta aquí bosquejados, los cuales permanecerán, en esencia, inalterados hasta el presente, es decir, durante unos 500 millones de años.

Todos los vertebrados somos cordados, y lo que define a un cordado es su notocorda, una estructura embrionaria que atraviesa longitudinalmente el cuerpo del embrión trazando su eje rostrocaudal, desempeñando funciones esenciales en el desarrollo del sistema nervioso (por ejemplo, en la inducción del tubo neural o, a nivel estructural, en el establecimiento de los patrones dorsoventral y mediolateral) y desapareciendo en la fase adulta (excepción hecha de los agnatos). Prolongándose a lo largo del eje que la notocorda prefigura encontramos en los vertebrados el sistema nervioso periférico, con una organización ganglionar similar a la del sistema nervioso de los invertebrados. A diferencia de lo que sucedía en éste, el sistema nervioso central de los vertebrados se encuentra, como apuntábamos, en una cavidad protegida por tejido óseo: el cráneo. Dentro de él ha venido buscándose solución al problema de la conciencia durante tres décadas. El diseño del sistema nervioso de los vertebrados se ha mantenido esencialmente invariable a lo largo de toda su historia evolutiva. Así, los vertebrados filogenéticamente más antiguos que comparten actualmente el planeta con nosotros, agnatos como la lamprea, presentan ya un encéfalo subdividido en tres regiones: el encéfalo ante-

rior (telencéfalo y diencéfalo), el medio (mesencéfalo) y el posterior (mielencéfalo y metencéfalo). Esta anatomía gruesa del sistema nervioso vertebrado, en contra de los postulados de populares aunque desacreditadas doctrinas como la del cerebro triúnico de Paul McLean, persistirá a lo largo de toda su filogenia y serán muy pocas las estructuras neuroanatómicas que a lo largo de la misma aparezcan diferencialmente en determinadas especies.[15] Las variaciones tendrán pues que ver principalmente con citoarquitecturas y tamaños relativos de núcleos, áreas o regiones, siendo la corteza cerebral la parte del telencéfalo más variable en este sentido –al punto que la expansión de los hemisferios cerebrales y, dentro de ellos, de las áreas de asociación, puede considerarse la marca distintiva de la evolución del sistema nervioso vertebrado y el sistema nervioso primate, respectivamente– y el tálamo la más variable del diencéfalo. Dada la implicación que según un abrumador cuerpo de evidencias ambas estructuras tienen en gran cantidad de experiencias conscientes, cabe especular acerca de una variabilidad fenomenológica interespecífica paralela a esta variabilidad anatómica.

Da en todo caso la impresión de que la historia evolutiva del sistema nervioso –sin la cual, sobra indicarlo, nuestra comprensión del mismo apenas excedería el mero inventariado: "a proper understanding of the brain can only really be achieved in the context of its evolution" (O'Shea, 2005: 49)– puede reconstruirse con mayor facilidad que la de la conciencia: los intermediarios teóricos de los que necesitamos servirnos para alcanzar y organizar los vestigios de la primera serán, con toda certeza, más acostumbrados y manejables y menos numerosos. Es, adicionalmente, más que probable que las especulaciones evolutivas que pretendan presentarse hoy como la solución –en singular– al problema de la conciencia acaben con el tiempo convirtiéndose en el estímulo o, en el mejor de los casos, la base para un fructífero abordaje de uno entre los muchos problemas (explicativos) en que el así llamado problema de la conciencia consiste. Y precisamente ahí, por otra parte, puede que resida su valor, que, apostemos, será relativo a la medida en que semejantes esfuerzos especulativos se encuentren acompañados del serio intento de aunar perspectivas dispersas en diferentes marcos teóricos y distintas áreas de investigación.

Una idea que, entendemos, ha venido obstaculizando nuestra comprensión de la evolución de la conciencia arraiga en la concepción cognitivista canónica según la cual

[15] Así, por ejemplo, el 95% de los núcleos que encontramos en el encéfalo de la rata pueden hallarse asimismo en el del ser humano (Miklos, 1998).

la mente es una especie de receptáculo del mundo. Esta mente receptáculo se encarga de elaborar modelos del mundo exterior de los que dimana, justamente, la utilidad de lo mental, su función. Desde esta perspectiva, la tarea que mantiene entretenido al telar mágico que los vertebrados llevamos incrustado en el cráneo es la urdimbre de modelos a escala del "mundo exterior" de los que, *ex hypothesi*, nos servimos para orientarnos y conducirnos en él. Si este urdir fuera el trabajo que la evolución hubiera tenido a bien asignar a lo mental, no podríamos llamar mental a nada que no requiriera del mismo, pero es ciertamente comprometido afirmar que el modo en que experimento el sabor del roquefort, el hambre o, incluso, el modo en que manipulo los adminículos de los que me sirvo para prepararme el desayuno precisen de semejante clase de modelos internos. De hecho, es más comprometido aún afirmar que ese urdir fuera el cometido de las regiones encefálicas que sustentaran la eclosión evolutiva de la experiencia consciente. Existen, como veremos, buenas razones biológicas para pensar que una idea bien alejada de esta concepción de lo mental tiene más posibilidades que la misma de conformar un sustrato fértil para la elaboración de un marco teórico en que incardinar la evolución de la conciencia. Podemos partir de las siguientes consideraciones hacia una exposición de la referida idea. Existe una clase muy específica de lesión neurológica capaz de abolir la conciencia: la integrada por aquéllas que afectan al sistema reticular de activación, cuyas funciones fueron atisbadas por Horace Winchell Magoun y Giuseppe Moruzzi en el último tramo de los cuarenta (Moruzzi & Magoun, 1949). El así llamado sistema reticular de activación ascendente es un heterogéneo y primitivo entramado neuronal que se extiende desde las protuberancias superiores del tronco del encéfalo y el mesencéfalo hasta el hipotálamo posterior, los núcleos intralaminares y reticulares y el cerebro basal anterior enviando eferencias difusas hacia el tálamo y la corteza. Su función ha venido interpretándose en términos de activación. El sistema reticular de activación marca el tono de la actividad talamocortical. Cuando este entramado neuronal dispara la activación del sistema talamocortical, promueve en el organismo un estado de alerta, cuando la inhibe, promueve uno de sopor, cuando deja de funcionar, ninguno (Plum, 1991). Desde las primeras intuiciones de Moruzzi y Magoun hasta nuestros días, el sistema reticular de activación ha venido estableciéndose como una estructura que juega un papel esencial en la orquesta de la conciencia (vid., v. g., Parvizi & Damasio, 2001). Este complejo entramado consiste, como señalábamos, en una gran cantidad de núcleos troncoencefálicos, diencefálicos y prosencefálicos basales que proyectan axones de largo trazado hacia el tálamo y la corteza interviniendo en la modulación de la actividad elec-

trofisiológica y neuroquímica de vastas regiones encefálicas, una modulación a la que ha venido denominándose arousal. En muy resumidas cuentas: alta activación talamocortical, alta responsividad, alta alerta; baja activación talamocortical, baja responsividad, baja alerta; activación talamocortical excesivamente alta o baja, responsividad y alerta excesivamente altas o bajas (éste es el esqueleto de la así llamada ley Yerkes-Dodson). A pesar de que actualmente sigue investigándose en humanos acerca de la relación entre arousal y vigilancia (vid. Sierra et al., 1993), operacionalizados mediante diferentes índices de activación fisiológica y subjetiva, y una enorme cantidad de procesos cognitivos, afectivos, motivacionales y conductuales, el protagonismo del arousal en la regulación del nivel de conciencia está hoy firmemente establecido –sin ir más lejos, por vía negativa, esto es, tomando en consideración los valores de los señalados índices durante el sueño de ondas lentas o el coma (vid. Laureys, Owen & Schiff, 2004) y cotejándolos con los mismos durante la vigilia, o cotejando entre sí los correspondientes a diferentes grados de alerta–. En la intersección entre el arousal y la emoción se encuentra la clave de la idea que avanzábamos. Los mecanismos neurofisiológicos responsables de ambos procesos son tan antiguos como el sistema nervioso de los vertebrados y, de acuerdo con la aparente tendencia de la selección natural a operar de forma más acusada sobre la periferia del sistema nervioso y favorecer una deriva evolutiva conservativa por lo que a su estructura básica respecta (Jing, Gillette & Weiss, 2009: 415), idénticos en lo esencial de los agnatos en adelante, pero antes que encontrarse diseñados para el corte y confección de representaciones internas de un mundo exterior que, por otra parte, se encuentra ya a disposición del organismo, lo están para cartografiar el estado de su cuerpo desde la perspectiva de sus necesidades biológicas primarias y propiciar respuestas integradas ante las disrupciones homeostáticas que constantemente sufre. La centralidad, el carácter imprescindible de la movilización del organismo de cara a evitar el equilibrio termodinámico ante las disrupciones homeostáticas que ponen en marcha la sinfonía fisiológica de las emociones y el arousal concede plausibilidad a la idea de que los primeros vestigios evolutivos de la conciencia surgieron no de la necesidad de elaborar un mapa interno del mundo exterior, sino de la de emprender cursos de acción integrados espoleados por estados emocionales básicos o elementales. Es interesante traer a colación en este punto el prestigioso manual *Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience*. Una de las escasísimas referencias que en el mismo encontramos a la conciencia aparece en el contexto de la introducción a la noción de homeostasis. En concreto, se nos dice (Rosenzweig, Breed-

love & Watson 1999/2005: 391 del original, 474 de la traducción) que los seres humanos sentimos determinadas disrupciones homeostáticas como frío, hambre o sed, y que a pesar de que los animales no humanos sean incapaces de informarnos lingüísticamente acerca de esas mismas experiencias, no parece excesivamente aventurado darlas por supuestas, ni tampoco suponer que un animal humano hambriento y un animal no humano en las mismas condiciones se sentirán, más o menos, del mismo modo.

La herencia intelectual de las tradiciones conductista y cognitivista, que vinieran definiendo lo mental en términos de estímulos, representaciones de estímulos y respuestas a estímulos, hace que siga resultando para muchos difícil adoptar una perspectiva acerca del origen evolutivo de un proceso mental en apariencia tan sofisticado como la conciencia en la cual los estímulos, las representaciones de los mismos, la manipulación de éstas y las respuestas a aquéllos juegan un papel secundario frente al de la emoción, una perspectiva que ubica la base de la vida consciente no en el fatigoso ejercicio de construir maquetas interiores del mundo exterior, sino en el proceso de cartografiar el estado del propio organismo desde el punto de vista de sus necesidades biológicas de cara a movilizar su actividad de acuerdo con las mismas. Al hablar aquí de emociones, por otra parte, no nos referimos en primera instancia al miedo, el asco o la sorpresa, sino, en línea con Denton (2005; Denton et al., 2009), al hambre, la sed, la sed de sal, la necesidad de excretar o la de respirar y, decisivamente, a sensaciones somáticas básicas como el dolor. Todas estas formas elementales de la vida afectiva, cuyo locus se halla en estructuras encefálicas de temprana aparición en el curso de la evolución y estrechamente ligadas a las dinámicas del estado del medio interno y la actividad del cuerpo del organismo al completo, con el que intercambian un bidireccional flujo masivo de señales, responden a necesidades biológicas fundamentales y se asocian a experiencias subjetivas específicas. Dichas experiencias se sitúan en el polo negativo de esa hipotética escala en cuyos extremos ubica el valor biológico lo agradable y lo desagradable: cuando las rupturas del equilibrio homeostático amenazan la vida del organismo, las mismas provocan sentimientos tanto más aversivos cuanto mayor sea la amenaza, sentimientos acompañados de estados de activación cuyo objeto es el de movilizar una rápida respuesta, que será asimismo tanto más vigorosa cuanto más apremiante sea la necesidad y que, en caso de éxito, esto es, en caso de servir al restablecimiento del señalado equilibrio, dará lugar a sentimientos agradables. Estas formas básicas de la vida afectiva, a causa de su vínculo con necesidades biológicas fundamentales, confieren al flujo de la experiencia su siempre presente tonalidad emocional, ocasionada por una prospección

primariamente troncoencefálica y constantemente actualizada del estado del organismo. Los resultados de esta prospección pueden preponderar sobre cualquier otro aspecto de nuestras vidas mentales, lo cual no es de extrañar, pues, ¿cabe experimentar algo por encima de la urgente ansia de aire cuando uno está ahogándose? Existen razones adicionales para el estatus especial de dichos resultados. Las regiones encefálicas que acotan el cuerpo en diferentes clases de mapas somáticos tienen con el mismo una relación directa y bidireccional, afectándose mutua y constantemente, sin intermediarios, una relación completamente diferente de la que mantienen con sus objetos las regiones encefálicas encargadas de mapear el mundo externo. Se trata de dos formas heterogéneas e incomparables de interacción (vid. Damasio, 2010: 89-90 del original, 148 de la traducción) a la vista de las cuales la idea de la precedencia de la afectividad en la filogénesis de la experiencia consciente resulta enteramente natural.

> Affective states generated by ancient brain regions may be the first kinds of experiences that existed on the face of the earth. Without them, consciousness may not have emerged in brain evolution (Panksepp, 2012: 322).

Estados emocionales como los ocasionados por estas necesidades biológicas fundamentales tienen su origen en un desequilibrio interno que exige la movilización del organismo para compensarlo, pero, en cualquier caso, lo que necesita un organismo en este sentido necesitado no es un mapa interno de su medio, sino enlazar apropiadamente en su interacción con su entorno cursos de acción con formas específicas de desequilibrio interno. Dada la centralidad de la afectividad y sus antiguos mecanismos fisiológicos en toda forma de experiencia, cabe conjeturar que en los orígenes evolutivos de la conciencia se hallara no la necesidad de representar o modelar el mundo exterior en un indefinido espacio interior, sino la de orquestar la actividad del organismo armonizando respuestas conductuales y necesidades biológicas. Como indicábamos, la concepción tradicional según la cual los ladrillos de lo mental son "agentes internos especializados en la recepción de la información disponible en la periferia del cuerpo" (Dennett, 1996a: 82 del original, 102 de la traducción) no podría distar más del lugar desde el que proponemos enfocar la panorámica a la que venimos apuntando, fundada en la idea de que "la conciencia surge dentro de la historia de la regulación biológica, que es un proceso dinámico conocido con el nombre de homeostasis" (Damasio, 2010: 25 del original, 52 de

la traducción),[16] y la de que los procesos emocionales básicos, "escudos protectores máximos de la supervivencia" (Mora Teruel, 2013) hasta el momento desatendidos por la mayoría de las teorías de la experiencia consciente, habitan la médula del fenómeno que dichas teorías debieran explicar y constituyen, de hecho, sus condiciones de posibilidad (Watt, 1998; 1999).

Si pudiéramos mantener nuestra homeostasis mirando el Sol, los animales seríamos plantas. Sin embargo, se da el caso de que debemos emprender acciones a tal fin. A ello se debe que tengamos sistema nervioso y nos veamos en la necesidad de desarrollar actividad cognitivo-afectiva y conductual. Si a dicha actividad le fuera dable cesar sin acarrear con ello nuestra muerte, bien podríamos hacer como los urocordados y comernos nuestro propio sistema nervioso llegado el momento de adoptar la quietud abandonando nuestro afán de desplazarnos de un lado a otro. Pero no, no le es dable y sólo cesará con la disolución del animal en el equilibrio termodinámico. El estímulo elemental de esta actividad reside, como sugeríamos, en la necesidad de mantener la homeostasis, pero no son las células sensibles a los estímulos externos las encargadas de poner en funcionamiento los mecanismos de alarma ante los desequilibrios de la misma, sino los interoceptores. La integración de las cascadas fisiológicas de interoceptores y exteroceptores se hace progresivamente más sofisticada durante la filogénesis gracias al surgimiento de estructuras encefálicas cada vez más complejas. De este modo, por ejemplo, los primeros tetrápodos podían morir deshidratados junto a un charco a no ser que cayeran en él por casualidad. Esto no habla necesariamente en contra de su sensación subjetiva de sed, sino de las bondades del ajuste de la integración que a su sistema nervioso le cabe realizar entre su sistema endógeno de necesidades y su sistema sensorial. Un reptil, sin embargo, correría un destino menos aciago dada su mayor capacidad para acoplar la actividad fisiológica que parte de los exteroceptores conformando el núcleo de la percepción con la que parte de los interoceptores conformando el de la emoción. Este acoplamiento entre el interior y el exterior motivado por las constantes fluctuaciones homeostáticas y mediado por una interacción entre sistemas interoceptivos y exteroceptivos destinada a disponer los medios para compensarlas podría ser la base sobre la que han venido alzándose los diferentes grados y formas de conciencia en vertebrados, una base cuyos elementos esenciales, la modulación del arousal y los estados emocionales más básicos, están encarnados en redes retroalimentadas, esencialmen-

[16] "The sentience that we experience as consciousness has its precursors in the adaptive processes of life in general" (Deacon, 2012: 22).

te troncoencefálicas, cuyo último propósito es el de guiar la conducta adaptativa del organismo habilitando la selección de respuestas óptimas a la vista de los resultados del constantemente actualizado proceso de mapeado del estado del cuerpo del organismo, pues, como un creciente cuerpo de evidencias pone de manifiesto, los mecanismos elementales del arousal y la emoción subyacen a todos los procesos de toma de decisión en vertebrados (Mashour & Alkire, 2013: 10359). "Las emociones y los sentimientos son el origen de la conducta (…), la energía que permite el ensamblaje coherente de todos los elementos de [la] planificación" (Mora Teruel, 2001: 106), "el motor que nos mantiene vivos" (Mora Teruel, 2004: 101), y, en este sentido, la emoción ha llegado a ser definida como un elemento esencial de toda forma de conducta dirigida a metas: "emotion is outward movement. It is the 'stretching forth' of intentionality, which is seen in primitive animals preparing to attack in order to gain food, territory, resources to reproduce, or to find shelter and escape impending harm" (Freeman, 2000a: 214). Esta hipótesis acerca del origen de la experiencia consciente encuentra respaldo en evidencia disponible acerca del papel esencial que los primitivos mecanismos fisiológicos del arousal y las formas elementales de la emoción juegan en el mantenimiento y modulación de la experiencia consciente en humanos, y en la esencial similitud entre las estructuras encefálicas que desencadenan las reacciones afectivas tanto de animales humanos como no humanos (Panksepp, 2011; Panksepp & Biven, 2012). Sin embargo, estas escuetas consideraciones han de leerse, exclusivamente, como el esbozo de una hipótesis plausible a la luz de la evidencia disponible, una hipótesis que es ya actualmente objeto de atención pero que precisa tanto de investigación empírica como, particularmente, de un tratamiento teórico capaz de integrarla en los sucesivos marcos de la fisiología comparada, la psicobiología comparada y la etología abriendo el camino a proyectos de investigación de creciente especificidad e interdisciplinariedad. ¿Podemos, no obstante, atisbar aquí el fermento de una teoría unificada de la evolución de la conciencia? A una pregunta como ésta cabe responder con otra. ¿Existe una teoría unificada de la mente? La respuesta es que no. Puede que este o aquel paradigma gane fuerza en esta o aquella subárea durante un determinado periodo de tiempo, pero lo más fructífero parece que seguirá siendo conducir investigaciones concretas dentro de subáreas particulares a la luz de diversos paradigmas y tratar al tiempo de perfilar tentativas teóricas destinadas a tender puentes entre esas subáreas y paradigmas. Las de «mente» y «conciencia» son voces que utilizamos para referirnos a una enorme multitud de procesos que muy probablemente sea recomendable seguir investigando dentro del marco de una considerable

cantidad de tradiciones teóricas y paradigmas experimentales, y lo mismo cabe decir de la indagación de su historia filogenética. El imperialismo y la uniformidad parecen, pues, las menos indicadas entre las opciones disponibles. La historia de las ciencias cognitivas, y particularmente la de la psicología, puede leerse como una guerra por el imperio, una guerra entre escuelas siempre fútil, a la postre: cada uno de los marcos teóricos que han presentado combate han legado a las sucesivas generaciones de investigadores herramientas teóricas, heurísticas y experimentales de las que éstas han sabido sacar un partido mayor o menor, pero con independencia de su alcurnia. La pluralidad no es sólo ineludible para una interdisciplina tan joven como los *Consciosuness Studies*, sino probablemente también lo más saludable. Se entiende, pues, que no pretendamos con esta tentativa hipótesis estar allanado *el* camino para *la* explicación de la evolución de la conciencia, entre otras cosas porque muy probablemente también carezca de sentido hablar en este punto en singular y una tal explicación se vea necesitada de un plural y multifronte trabajo experimental y teórico capaz de simultanear y articular diferentes perspectivas y niveles. Las palabras de Endel Tulvin en su introducción a la sección sobre la memoria de la segunda edición de *The New Cognitive Neurosciences* nos vienen aquí como anillo al dedo –sólo debemos obviar que el estonio-canadiense se refería a la memoria en lugar de a la conciencia.

> It is a ubiquitous presence in all higher life forms. It take many shapes, from simple to complex, from highly specific to most general, from trifling to fundamentally important. In its manifold expressions it is being observed [and] investigated (...) in numerous organisms, at many levels of analysis from a variety of vantage points and relying on many different approaches and techniques (Tulving, 2000: 727).

Nuestra hipótesis acerca del origen afectivo de la experiencia consciente, fundada en la centralidad de los mecanismos emocionales, ha de entenderse como una alternativa evolucionista no excluyente, sino, sencillamente, más básica que las ofrecidas desde el prisma funcionalista, fundadas en la centralidad de los mecanismos perceptivos y cognitivos. En este sentido, incluso la teoría de Edelman –que no sólo se muestra como una de las más robustas y completas sino que, adicionalmente, "ofrece un prometedor marco biológico de investigación" (Cavanna & Nani, 2014: 131τ)– sitúa el origen de la experiencia, la conciencia primaria, en el vínculo adaptativo y dependiente del valor aportado por la neuromodulación dopaminérgica, colinérgica y noradrenérgica tendido entre las actuales categorizaciones perceptuales y las respuestas de aprendizaje prove-

nientes de la historia de relaciones del organismo con su entorno.[17] Este énfasis en el aprendizaje y, particularmente, en la percepción aparece de forma más explícita en las teorías de Zeki, Lamme, Milner & Goodale y Crick & Koch, que de hecho no toman en consideración ninguna forma de experiencia consciente al margen de las perceptivas, pero se encuentra presente en las de Gazzaniga, Ramachandran, Dehaene, Naccache & Changeux y, en verdad, en todas las teorías de la conciencia excepto en la de Damasio y, en menor medida, la de Llinás, que concibe la evolución de la mente como una progresiva internalización de la motilidad en la que la subjetividad emerge a expensas de la progresivamente ampliada capacidad predictiva que un sistema nervioso cada vez más complejo propicia.

Nuestra hipótesis acerca de la filogénesis de la conciencia excluye toda apelación a formas complejas de cognición como función u origen de la experiencia consciente, y en particular toda homologación entre experiencia consciente y capacidad para informar de la misma a través del lenguaje humano. En los inicios de la investigación y la teorización científica acerca de la conciencia ésta fue situada en la cúspide del sistema cognitivo y a menudo identificada con formas de procesamiento de la información extremadamente complejas. A pesar del declive de esta tendencia, sigue resultando habitual que la conciencia sea presentada como uno de los más sofisticados y elevados segmentos del sistema cognitivo, y en muchas ocasiones como un segmento dependiente de nuestra capacidad lingüística, como sucede, por ejemplo, en la teoría de Ramachandran o en la de Gazzaniga. Quizá el mejor antídoto contra esta clase de identificación de la conciencia con la capacidad lingüística y, particularmente, contra la idea de que el lenguaje juega algún papel en el origen de la experiencia consciente sean los resultados experimentales obtenidos en estudios de discriminación de drogas (Díaz & Velázquez, 2000), unos resultados que ponen de manifiesto que las sustancias psicoactivas que pro-

[17] Bien es cierto que Edelman (2003), tal y como Denton hace notar, incorpora *en cierta medida* (Denton, 2005: 105 del original, 165 de la traducción) la centralidad de los primitivos mecanismos emocionales y del control homeostático presentándolos como unos de los "primeros y más persistentes" (Edelman, 2003: 5523τ) insumos al núcleo dinámico, pero también lo es que en su último libro sobre la conciencia, publicado diez meses después que el artículo que acabamos de citar, Edelman vuelve a situar el origen evolutivo de la conciencia en el tránsito de reptiles a aves y mamíferos por cuanto sostiene que fue entonces que surgiera una nueva conectividad recíproca en el sistema talamocortical, en concreto, entre las áreas corticales encargadas de la categorización perceptiva y las responsables de la memoria valoral-categorial. La interacción dinámica entre memoria y percepción sigue cimentando el flanco histórico de la teoría de Edelman mientras la afectividad juega un papel secundario en la misma (vid., especialmente, Edelman, 2004: 54-55). La perspectiva de Edelman en sus últimas publicaciones se mantiene, pues, en la línea que trazara ya en su primera aproximación a la conciencia, en la cual la vincula con las funciones cerebrales superiores y las capacidades cognitivas complejas (Edelman, 1978: 51), presentándola como producto de la señalización de reentrada entre procesos paralelos que implican asociaciones entre patrones mnésicos y perceptivos (Ibíd.: 95).

ducen efectos subjetivos similares en humanos lo hacen asimismo en animales, siendo así que "los estudios de discriminación de drogas son los mejores modelos animales actualmente disponibles para estimar los efectos subjetivos de las drogas en humanos [y ello, obviamente, a causa de la] capacidad de los animales para discriminar los efectos subjetivos de las drogas" (Ambrosio Flores, 2004: 33-34). Por su parte, cabe basar argumentos a favor del origen troncoencefálico de la conciencia y contra la identificación de ésta con formas complejas de cognición, y, particularmente, contra la idea de que las mismas juegan algún papel en la emergencia evolutiva de la experiencia consciente, en la rara condición anatómica conocida como hidranencefalia, un desorden de etiopatogenia infecciosa o traumática cuya principal característica consiste en que, con frecuencia, quienes la padecen conservan sólo intactas las estructuras troncoencefálicas, apareciendo ausentes o gravemente afectadas todas las ubicadas por encima del mesencéfalo. No obstante, los niños hidranencefálicos muestran una gama de respuestas conductuales y expresivas incompatibles con la ausencia de experiencia consciente (Aleman & Merker, 2014). Desde luego, no son capaces de ninguna forma superior de cognición, pero muestran preferencia por estímulos visuales, cuidadores, sonidos e incluso melodías, manifiestan agrado o displacer, encontrándose su expresividad emocional en buena medida preservada, y pueden, de hecho, desplazarse hacia fuentes de sensaciones placenteras, por ejemplo, gateando hacia el lugar en que incide la luz solar (Damasio, 2010: 80 del original, 135 de la traducción).[18]

Contraponer al énfasis intelectualista en la percepción y los mecanismos cognitivos superiores la necesidad de ubicar en primer plano la vida afectiva abre para la biología de la conciencia posibilidades teóricas difícilmente explorables desde de la atalaya intelectualista cartesiana, la principal de entre las cuales es la de elaborar esquemas filogenéticos en los que la experiencia consciente aparece en continuidad con los mecanismos del control homeostático y la regulación de la vida, esquemas filogenéticos capaces de hacer justicia al modo en que toda lógica evolutiva apunta a la inquebrantable unión de la experiencia consciente con el valor biológico, la cual, desde luego, aparece desfigurada si posponemos nuestra exploración de la filogénesis del sistema nervioso en

[18] Hasta donde sabemos, Alan Shewmon fue el primero en apoyarse en la hidranencefalia para atacar la "doctrina cortical de la conciencia" (Shewmon, 1997: 58τ). Más recientemente, dicha doctrina ha sido asimismo desafiada desde una aproximación experimental en estudios de neuroimagen sobre la actividad neurofisiológica durante el despertar de la anestesia general (Långsjö, et al., 2012) y el sueño de ondas lentas (Balkin, et al., 2002). Esta doctrina aparece por otra parte ligada en el estudio neurocientífico de la emoción a las dicotomías cortical/subcortical, cognición/emoción, explícito/implícito y consciente/inconsciente –lo cual puede apreciarse particularmente bien en exposiciones de carácter didáctico (vid., v. g., Enríquez de Valenzuela, 2014c).

busca del valor biológico de la conciencia hasta la irrupción de dispositivos cognitivos tan sofisticados como el lenguaje.

La conciencia de los animales no humanos es todavía para muchos un tema tabú. La concepción de la evolución de la conciencia que de forma esquemática y tentativa avanzamos en este apartado se alza sobre el supuesto de que puede hablarse, con pleno sentido, de estados mentales conscientes en animales no humanos. Desafiar dicho tabú y acreditar el pleno sentido de dicha locución fue el objetivo que se marcara el grupo de neurocientíficos que el día siete de julio de 2012 firmara, durante la *Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals*, en una ceremonia en la que Stephen Hawking fue el invitado de honor, la *Cambridge Declaration on Consciousness*. Dicho documento no pretende avalar de forma exclusiva su desafío al apuntado tabú mediante pruebas tipo Gallup o Hans el listo, es decir, no trata mover hacia la aquiescencia respecto de la pertinencia de la atribución de conciencia a animales no humanos apelando únicamente a la autoconciencia o la capacidad para realizar operaciones cognitivas complejas, sino que recurre a tal fin a la neurofisiología de la emoción y las conductas emocionales básicas. Sin embargo, hasta el momento, todas las que se han presentado como pruebas de conciencia en animales no humanos han venido de hecho consistiendo en pruebas de habilidades cognitivas complejas. Y, ciertamente, existe una enorme cantidad de experimentos cuidadosamente diseñados que hacen patente el despliegue por parte de animales no humanos de capacidades cognitivas que difícilmente cabe interpretar como desligadas de cualquier clase de experiencia consciente y que vendrían a ejemplificar un complejo entreveramiento entre metas, planes y conducta. Así, por ejemplo, basándose en los resultados obtenidos en una serie de experimentos que ha sido ampliamente comentada (Clayton & Dickinson, 1998; Clayton et al., 2000), el equipo de la psicóloga británica Nicola Susan Clayton ha venido sosteniendo la existencia en aves de una forma de memoria que en humanos se concibe como necesariamente vinculada con la experiencia consciente: la memoria episódica (Tulving, 1972). Este tipo de memoria, encargado de codificar y recuperar información acerca de qué, dónde y cuándo tuvo lugar un suceso dado, a pesar de las pruebas consistentes acerca de la pertinencia de la adscripción, viene atribuyéndose a animales no humanos con la mayor cautela y denominándose en su caso "memoria *tipo* episódica" (episodic-like memory) dado que sigue discutiéndose si la misma se encuentra acompañada de experiencia consciente, atributo definitorio de esta clase de memoria en humanos, como indicábamos. Desde que el equipo de Clayton publicara sus resultados poniendo sobre

la mesa por vez primera la capacidad de un no primate –en concreto, un córvido: *aphelocoma californica*– de imaginar el futuro, planificar y actuar en consecuencia, experimentos análogos han venido arrojando evidencias en el mismo sentido en diferentes especies de vertebrados, pero también en invertebrados (vid., v. g., Menzel, 2009). No obstante, si la conciencia se halla diseminada, como proponemos, a lo largo y ancho del reino animal en diferentes grados y variedades, el criterio mínimo para la adscripción de la misma no tendría por qué apelar a algo tan sofisticado como la memoria episódica. Pero la cuestión es que no existe hoy por hoy acuerdo acerca de los criterios conductuales en los que fundar de forma fidedigna nuestras atribuciones de conciencia a animales no humanos, bien pretendan las mismas asentarse sobre la memoria episódica o sobre cualquier capacidad cognitiva más básica. Quizá una buena idea en este sentido sea, como hemos venido sugiriendo, continuar alejándonos de la tradición intelectualista cartesiana, a cuyo socaire vino identificándose la mente con la cognición, y comenzar a prestar atención a la mente afectiva. Con todo, reconstruir de forma cabal la filogénesis de la conciencia es un proyecto en cuyo éxito cabe depositar escasas esperanzas, al menos en el corto plazo. Pasarán décadas y el taraceado de este flanco explicativo seguirá repleto de huecos vacíos o rellenos de especulaciones que inicialmente puedan antojársenos inconfutables y desacuerdos que inicialmente puedan antojársenos inconciliables. En cualquier caso, arrostrar la tarea de elevar el edificio de este flanco explicativo brindará sin duda mayor retribución que condenarlo a priori al desmérito y el fracaso, pero ello no será posible si no logramos antes vencer la injustificada tentación del escepticismo acerca de la conciencia en animales no humanos (Cartmill, 2000: 845).

### 2.2._La función de la experiencia II. Qué hace la conciencia

Si debemos concebir a la experiencia consciente como un rasgo biológico, desentrañar la función que desempeña consistirá en encontrar su *valor causal*, es decir, aislar de la plétora de relaciones causales en la que, como todo rasgo biológico, se encuentra implicada, aquéllas en las que reside eso que, en un sentido biológicamente relevante, la misma hace. Hablar de sentidos biológicamente relevantes significa aquí, por una parte, que si bien el páncreas tiene una enorme cantidad de poderes causales, de los que hablamos cuando nos referimos a su función biológica son aquéllos en virtud de los cuales contribuye a mantener temporalmente al organismo alejado del equilibrio termodinámico y, por otra, que si bien puede decirse que su función exocrina es la liberación de en-

cimas digestivas en el intestino delgado, esta función no cobra sentido sino en el marco de su integración con la actividad convergente de otros órganos y sistemas: en último término, podría decirse, las funciones exocrinas y endocrinas del páncreas estriban en la captación, transferencia y liberación de la energía contenida en los enlaces que unen los grupos fosfato con el resto de la molécula de ATP, pero esto, obviamente, requiere de la contribución del sistema digestivo y circulatorio al completo, así como de la de diversos orgánulos celulares. Hablar de funciones biológicas, pues, no es sino hablar de potencia causal, esto es, de la capacidad de un determinado rasgo para producir determinados efectos, pero *desde un determinado punto de vista*. El corazón bombea sangre y ésa es su función, pero dicha función sólo cobra sentido cuando es contemplada en un contexto mayor (el del oxígeno bañando los diferentes tejidos, etc.). Además, al desempeñar su función, produce otros efectos irrelevantes desde los puntos de vista biológico y explicativo (como el sonido que producen sus latidos), con lo cual se hace enteramente claro que la noción de función biológica corresponde a la de la potencia causal de un rasgo contemplada desde un determinado punto de vista, el de sus efectos relevantes por lo que a la provisoria prolongación de la vida del organismo se refiere y, así, relevantes por lo que a la explicación biológica toca.

No vamos a abrir una caja y hallar en ella la función de la experiencia consciente, dado que la misma desempeña, con toda probabilidad, diferentes funciones analizables en diferentes niveles de complejidad e integradas en la dinámica de la interacción entre el organismo y su medio. Se han propuesto una gran cantidad de posibles funciones de la experiencia consciente. El trabajo que queda por hacer es el de analizar el alcance, dominio e idoneidad de las mismas integrándolas en el marco de los pertinentes resultados experimentales y herramientas teóricas procedentes de diversas áreas de investigación proponiendo, adicionalmente, vías para su operacionalización y su tratamiento experimental.

En una época tan temprana como 1972 el psicólogo estadounidense Robert Evan Ornstein, en su libro *The Psychology of Consciousness*, propuso una serie de funciones de la conciencia que siguen, como veremos, presentes en el debate contemporáneo. En concreto, sugirió que la conciencia a) simplifica y selecciona información de entre la enorme cantidad presente en cada momento en nuestras mentes-cerebros, b) guía u orienta y supervisa la acción del organismo coordinando estados del cuerpo, el sistema nervioso y el entorno, estableciendo además prioridades para la acción al poner en relación el orden de necesidades biológicas con la señalada coordinación, y c) detecta y

resuelve discrepancias poniéndonos antes al corriente de lo inusual que de lo acostumbrado (Ornstein, 1972/1986: 64). Presentaremos brevemente a continuación las tres funciones de la conciencia que actualmente gozan de mayor aceptación. El modo y el contexto en que se presentan, así como su operacionalización, divergen en buena medida de la formulación pionera de Ornstein, pero comprobaremos que las funciones que en la misma atribuía a la conciencia pueden entenderse como comprendidas, respectivamente, en las tres siguientes, ésas a las que nos referíamos como las más aceptadas en la actualidad.

a) *Discriminación*

La idea de que la función de la conciencia pueda residir en el modo en que la misma se halla vinculada con la discriminación ha sido minuciosamente elaborada y defendida por Edelman y Tononi (Tononi & Edelman, 1998; Edelman & Tononi, 2000). En su marco teórico, la noción de discriminación presenta a cada experiencia consciente como única –en tanto su aparición excluye la de una enorme cantidad de experiencias alternativas– y se encuentra integrada en la teoría matemática de la información, quedando definida como la informatividad de una experiencia consciente en virtud de la reducción de la incertidumbre que propicia. La idea, en muy resumidas cuentas, sería que cada estado consciente es seleccionado en una fracción de segundo de entre una enorme cantidad de estados conscientes posibles, cada uno con diversas conexiones potenciales con la conducta del organismo y, así, lo que hace que un estado consciente sea informativo no tiene nada que ver desde este punto de vista con el modo en que pueda concebirse como compuesto por bits o pedacitos de información, sino con el hecho de que, al tener lugar, discrimina entre billones de estados posibles, cada uno de los cuales podría intervenir diferencialmente en la economía conductual del organismo. La informatividad se encuentra así en el núcleo de esta propuesta, pues sería la misma la que resultaría adaptativa: pudiendo cada experiencia consciente diferenciada vincularse con respuestas conductuales asimismo diferenciadas, la función de la experiencia sería la de aumentar la discriminatividad y, por tanto, la flexibilidad y adaptabilidad de la conducta. Como se desprende de lo señalado en el capítulo quinto de la primera parte, la noción de discriminación de la que aquí hablamos halla acomodo en la hipótesis del núcleo dinámico, que relaciona la discriminatividad de una experiencia consciente dada con la de la dinámica neuronal que la sustenta, la cual puede ponderarse mediante su

grado de complejidad neuronal, una medida basada en la teoría de la información que expresa el grado en que grandes subconjuntos de un sistema tienden a comportarse de manera coherente mientras pequeños subconjuntos tienden a hacerlo de forma independiente. La hipótesis de Edelman y Tononi es de gran feracidad heurística, apela a mecanismos neuronales concretos y definidos e interpreta la funcionalidad de la experiencia en términos operacionalizables y suficientemente básicos y flexibles como para permitir su expansión desde el contexto perceptualista en que fuera formulada hacia el afectivo –en cuya conveniencia hemos venido incidiendo–. No obstante, y a pesar de las posibilidades que abre su apelación a la complejidad neuronal, la misma ha de ser tratada con cautela, dado que interpretar esta medida como el rasgo definitorio de la experiencia consciente, tal y como ha venido haciendo Tononi en su reelaboración de la misma en términos de integración de información, puede dar lugar a dificultades teóricas tales como la posibilidad de atribuir conciencia a sistemas inconscientes por el mero hecho de que los mismos puedan conceptualizarse como poseedores de elevados niveles de integración o complejidad en el sentido propuesto en la hipótesis.

b) *Integración y flexibilidad*

Vinculada con la anterior encontramos la idea de que la función de la conciencia podría residir en la integración multimodal de actividad psicofisiológica de otro modo independiente que, dada precisamente dicha integración, abriría la puerta a formas de conducta flexibles y dependientes del contexto. La teoría neurobiológica de Edelman y Tononi y la cognitiva de Baars ofrecen diferentes conceptualizaciones de esta posible función de la experiencia consciente. Por lo que a esta última respecta, ya el propio utillaje cognitivo utilizado en su formulación dificulta su extensión aproblemática desde el tratamiento de la conciencia en términos de conciencia de acceso a la de la experiencia en términos afectivo-cognitivos. En cualquier caso, la teoría presenta como conscientes aquellos contenidos capaces de acceder al espacio de trabajo global, y a éste como un recurso central que permite la integración de la actividad cognitiva que de otro modo tendría lugar en procesadores especializados y funcionalmente aislados. Desde su primera formulación, la teoría del espacio de trabajo global apuntaba a una concepción de la función de la conciencia fuertemente apegada a los marcos teóricos y evidencias experimentales obtenidas en psicología cognitiva. Tanto éstas como aquéllos venían presentando a los procesos inconscientes como rápidos, limitados en capacidad e inflexi-

bles y a los conscientes como lentos pero en buena medida exentos de la referida limitación y mucho más flexibles, y precisamente en esta flexibilidad residiría la función de la conciencia, pues mientras que en situaciones comunes las respuestas conductuales inconscientes pueden dispararse como rutinas automatizadas de forma rápida y eficaz, cuando el organismo se enfrenta a situaciones nuevas en las que esos automatismos no resultarían adaptativos, la integración multimodal y la difusión de información pertinente a través de procesadores de otro modo aislados serían las instancias que habilitarían la flexibilidad necesaria para la producción de respuestas conductuales novedosas. Partidarios de la teoría del espacio de trabajo global neuronal han ampliado el planteamiento original de Baars para defender la idea de que la clase de integración de procesos cognitivos a la que la misma alude permitiría formas de simulación mental y evaluación interna en virtud de las cuales la selección de cursos de acción tendría lugar de una forma menos rígida –en el sentido de la indeterminación contextual– y más económica –en el sentido de requerir una menor cantidad de actividad biológica–, abriendo el camino de una progresiva internalización representacional del entorno del organismo y las respuestas de éste en el mismo, una suerte de virtualización que permitiría al organismo ir más allá de las configuraciones estimulares inmediatas mediante la simulación mental de su entorno y su conducta.

Por su parte, en la teoría de Edelman y Tononi la flexibilidad conductual aparece como un sólido candidato al título de valor adaptativo de la experiencia consciente y como derivada de la comentada discriminatividad –que los autores vinculan asimismo con la accesibilidad global (Edelman & Tononi, 2000: 147)– y del carácter dinámico de la integración neuronal en que la misma se basa. Dicha teoría, a diferencia de la del espacio de trabajo global neuronal, de cuño marcadamente cognitivista, abre vías para un estudio experimental de la función de la experiencia acorde con los principios que hemos propuesto para una biología de la conciencia y capaz de relacionar la integración neuronal y la flexibilidad conductual con la discriminatividad experiencial al vincular dinámicas conductuales y neuronales con presentaciones estimulares y contextos experimentales diversos desde el punto de vista de la experiencia (enmascaramiento, diferentes niveles de arousal, etc.). Elucidar qué es lo que la conciencia hace pasa por un cuidadoso cotejo de la actividad del organismo en diferentes estados, desde la vigilia con apropiados niveles de arousal hasta estados de ausencia o grave alteración de la conciencia, tales como los propios de determinadas clases de crisis epiléptica (crisis de ausencia, poriomanía epiléptica), sonambulismo o intoxicación. Los datos disponibles a

día de hoy, procedentes de estudios realizados con sujetos humanos (vid., v. g., Cavanna, 2008; Plazzi et al., 2005), apuntan a una misma conclusión: en cada una de las situaciones señaladas el repertorio conductual se reduce drásticamente y el comportamiento de los sujetos se caracteriza por la rigidez de las estereotipias y los movimientos repetitivos sin objeto aparente. Por otra parte, está bien establecido en la literatura neuropsicológica que enfrentarse a problemas nuevos y complejos requiere de una atención consciente que, con la repetición de la solución hallada (piénsese en interpretar una nueva pieza al piano o en aprender a conducir), se atenúa hasta desvanecerse con la conversión de dicha solución en una rutina automatizada, un tránsito desde la atención consciente a la rutina acompañado por un descenso de la actividad cortical a los ganglios basales. Tanto esta perspectiva de la automatización como la anterior de la restricción conductual en estados de ausencia o grave alteración de la conciencia podrían integrarse adecuadamente en paradigmas experimentales cuidadosamente diseñados y justificados de cara a ofrecer soporte empírico a hipótesis capaces de trazar puentes entre la flexibilidad conductual y las nociones de integración y discriminación tal y como Edelman y sus colaboradores han venido operacionalizándolas.

c) *Predicción*

En concordancia con hipótesis provenientes de implementaciones neurobiológicas de la teoría cognitiva del espacio de trabajo global según las cuales una tal arquitectura habilita para la simulación mental y la evaluación interna de acciones no llevadas efectivamente a término, con la evidente ventaja de ahorrarle al organismo la energía y el riesgo de realizarlas (Dehaene & Naccache, 2001),[19] encontramos la de Jeffrey Alan Gray, según la cual aquello que experimentamos conscientemente es seleccionado por su valor en una escala graduada entre lo esperado y lo inesperado (Gray, 2004: 232). Esta propuesta, asentada sobre una concepción del cerebro entendido como un sistema de comparación que predice lo que debe suceder y detecta desviaciones sobre esa predicción, vendría a ofrecer explicación al hecho de que durante la realización de rutinas automatizadas no experimentamos conscientemente detalles relacionados con el contexto en que las ejecutamos ni con nuestra ejecución de las mismas, a no ser que surja un

[19] Revonsuo (2005), en una línea similar, y después de haber defendido que su aproximación a la función biológica de las ensoñaciones desde su teoría de la simulación de la amenaza (Revonsuo, 2000) no deja de tener consecuencias para la teorización acerca de la función de la experiencia consciente (Revonsuo & Valli, 2000), sitúa a ésta en la simulación interna de la conducta y la percepción.

imprevisto, con lo cual la función de la conciencia es presentada como la de permitir corregir errores en las inferencias y predicciones inconscientes que constantemente genera nuestro sistema nervioso durante el desarrollo de nuestra conducta.

Las funciones de la conciencia que hemos comentado –así como otras referidas a procesos mentales de nivel superior, como la de la inferencia de los contenidos de otras mentes, relacionada con el carácter social de nuestra especie (vid. Humphrey, 1987)– resultan plausibles y ofrecen perspectivas fructíferas para su estudio experimental. Por otra parte, todas ellas pueden resumirse en una sola, la de posibilitar pautas adaptativas de conducta cuando los automatismos se muestran insuficientes. Sin embargo, hasta el día de hoy la función de la conciencia se ha buscado partiendo de teorías que presentan el origen evolutivo de la experiencia como esencialmente relacionado con la actividad perceptiva, la conformación de representaciones del mundo externo y el funcionamiento de los segmentos superiores del aparato cognitivo. Nosotros hemos defendido la pertinencia de buscar ese origen no exclusivamente en la relación perceptiva y cognitiva con el entorno, sino asimismo, y primariamente, en el soporte y motor de la misma en el mundo interno de las emociones primordiales producidas por las necesidades biológicas básicas. En este sentido, las funciones de la conciencia que hemos comentado pueden vincularse con este origen emocional y motivacional, pero los marcos teóricos desde los que han sido propuestas han impuesto una unilateralidad cognitivista y perceptualista a su formulación que no ha de implicar necesariamente, como sugeríamos, la imposibilidad de integrarlas en un marco teórico extendido que las articule con solvencia con la mente afectiva.

La función primigenia de la experiencia será con toda seguridad la función de la experiencia primigenia: formas sucesivamente más sofisticadas y complejas de experiencia pueden compartir algunos aspectos funcionales con las más básicas, pero la idea de que la experiencia cumple una misma función biológica a lo largo de toda la gama de clases de experiencia resulta un tanto extravagante y la alternativa ecléctica vuelve a presentarse como más prometedora. De este modo, parece que de poco valdrá discutir cuál es (en singular) la función de la conciencia cuando sigue abierta la posibilidad de estudiar las funciones de la experiencia consciente en contextos experimentales que cubren el espectro completo de lo mental, desde la emoción, la atención, la percepción, la planificación o la toma de decisiones hasta las reacciones de estrés en paradigmas psiconeuroinmunológicos. No es necesaria otra guerra interparadigmática: las hipótesis

acerca de la función de la conciencia pueden no sólo competir, sino también complementarse y dar cuenta, solapada o diferencialmente, de fenómenos ubicados en diferentes niveles de lo mental. Pero, en cualquier caso, ¿qué cabe entender que hacían y hacen las formas de experiencia que hemos presentado como básicas? ¿Cómo pudo intervenir en procesos selectivos? ¿Cuál es su valor biológico? En otras palabras, y en resumen, ¿cuál su función? Algunos (vid., v. g., Dickinson & Balleine, 2000) han buscado este papel primordial de la experiencia en la relación entre el sistema motivacional, que arraiga en las necesidades biológicas de los organismos, y el sistema cognitivo, que controla la acción dirigida a metas de los mismos. Aun cuando la división obedezca a fines expositivos (Damasio, 1994), y aun cuando no pocos se mostrarían renuentes a hablar de cognición en referencia a formas simples de conducta espoleadas por necesidades biológicas primordiales, los datos experimentales y la lógica evolutiva, como hemos tratado de mostrar a lo largo de este capítulo, presentan el proyecto de rehabilitar la centralidad de los hasta ahora en este punto desatendidos márgenes afectivos y motivacionales de lo mental como enteramente justificado, en una línea que conduce a definir la función primordial de la experiencia consciente como la de priorizar y promover cursos de acción relativos a necesidades biológicas en curso.

# III

# PARTE TERCERA

## Crítica de los principales marcos teóricos

# CAPÍTULO 10

## EL "NATURALISMO" "BIOLÓGICO" DE SEARLE

### 1._El marco filosófico de la aproximación de Searle al problema de la conciencia

Una anotación preliminar acerca del modo en que Searle aborda los extremos filosóficos de los que ha venido ocupándose nos pondrá en la pista del tono de su producción filosófica relacionada con el problema de la conciencia. El estilo de Searle es habitualmente –en contra de la norma en filosofía de la mente– poco técnico. Daniel Dennett ha llegado a referirse a la impresión causada por la literatura filosófica searleana en términos de "característico sabor populista" (Dennett, 1993a: 193τ). Puede que, en función del oído a ella prestado, la frase suene peor o mejor, pero lo cierto es que en ocasiones Searle parece obviar la complejidad de las discusiones en que sus colegas se hallan envueltos y simplifica de forma excesiva las perspectivas de los filósofos que critica –y Dennett, ciertamente, es una de las dianas que más ansían sus dardos–.[1] Acerca del estilo de Searle hemos dicho ya quizá más de lo necesario. Respecto de su forma de abordar los extremos filosóficos a los que ha venido dedicándose, Searle se refiere al análisis lógico (vid., v. g., Faigenbaum, 2003: 181-182) propio de la filosofía de corte analítico como su base metodológica y, en este sentido (vid., v. g., Searle, 1997b: 8; o el tercer párrafo de la introducción a Searle, 2002a) ha incidido en que su forma de abordar problemas filosóficos consiste en atender en primer término a *la pregunta* y tratar de ofre-

[1] Dennett (1993a: 204) hipotetiza que Searle, imbuido por una ofuscadora confianza en sí mismo, considera al referirse a otros autores las propuestas de éstos tal y como él las recuerda por lecturas que, en algún momento, realizó atentamente, pero que debería, en cualquier caso, refrescar, dado que parece haber olvidado los detalles más sutiles.

cer respuestas para la misma sólo después de haberla analizado, lo cual lleva en muchas ocasiones, propone, al desmantelamiento de problemas filosóficos, dado que permite advertir si aquélla o alguno de sus elementos se hallaban sustentados por supuestos erróneos. No obstante, su realismo, pretendidamente ajeno a lastres propios del anclaje en tradiciones filosóficas particulares, le lleva a complementar esta aproximación lingüística a los problemas filosóficos –que hereda de algunas de las figuras de la historia de la filosofía que más directo influjo han ejercido en su trayectoria, como Wittgenstein o Austin[2]– con una atención, teñida de cierta suerte de cientismo, a lo que –desde lo el propio Searle (Searle, 2004a) ha tenido a bien llamar un *realismo ingenuo* (*naïve*)– denomina *los hechos*[3] (ciertamente no en el sentido fenomenológico de *las cosas mismas*). Así, invita a "olvidar las grandes categorías y tratar de describir los hechos" (Searle, 2004a: 125τ) y prescribe: "cualquier teoría filosófica ha de ser concordante con los hechos" (Searle, 2007a: 325τ).

Este realismo de Searle es, en primer lugar, un *realismo externo* –que Searle denominó inicialmente *realismo metafísico* (Searle, 1991a)–, según el cual, y tal y como ha subrayado (en los primeros compases de Searle, 1997b), existe *un mundo real* con total independencia de nuestras representaciones, experiencias, opiniones, juicios o sentimientos. Es este aspecto del realismo, que en Searle funciona como una presuposición de trasfondo –"background of our thought and language" (Searle, 1997b: 9), trasfondo que Searle propone como anterior incluso a cualquier opinión de sentido común–, el que le ha llevado a defender tanto la concepción tradicional (de Aristóteles a Tarski) de la verdad como correspondencia (Searle, 1995a) como la versión mentalista del realismo ingenuo que encontramos a la base de sus planteamientos acerca del problema de la conciencia.

Además de la tesis ontológica del realismo externo, el realismo de Searle se presenta en una forma epistemológica dura. En primer lugar, su realismo ingenuo es una tesis que atañe a la filosofía de la percepción, por cuanto Searle ha argumentado que en la percepción se nos presentan directamente los objetos tal y como son (Searle, 1991a: 189).[4] Por otra parte, en la vertiente decididamente epistemológica de este realismo, el filósofo de Denver ha propuesto que podemos agradecer a nuestras capacidades cogniti-

[2] Bajo cuya dirección trabajara Searle durante los cincuenta en Oxford.
[3] Por los cuales, y según sus propias palabras, adquiriera en sus años de estudiante en Oxford verdadero respeto (vid. Faigenbaum, 2003: 21).
[4] Zemach (1991: 171) ha propuesto en este punto que, a falta de argumentos sólidos en defensa de dicha tesis, Searle debiera hablar más bien de la misma en términos de asunción.

vas la posibilidad de conocer el mundo real –lo cual no implica en su opinión que tales capacidades no sean limitadas, sino que, meramente, no son falaces–.[5] Desde este punto de vista ha argumentado que tal vez el escepticismo sea irrefutable, que no podemos probar la verdad del realismo ingenuo, sino la ininteligibilidad de su negación dentro del marco de un lenguaje público (Searle, 2004a: 277) sustentado por el acceso común y compartido a una y la misma realidad (vid., asimismo, el penúltimo capítulo de Searle, 1995a). Esta idea searlena de un acceso fiable a la realidad huye, sin embargo, de una concepción objetivista desde la perspectiva de la cual cabría interpretar, desde el punto de vista ontológico, que la realidad es total o exclusivamente objetiva, dado que, precisamente, Searle está particularmente interesado en presentar a la subjetividad como un elemento o componente irreductible y de pleno derecho de la realidad. En su planteamiento, tan inconcuso resulta el hecho de que existe un mundo independiente de nuestras experiencias como el de que existe el mundo de nuestras experiencias, es decir, lo que él denomina subjetividad ontológica: "*ontological* subjectivity and objectivity are [different but both real] features of *reality*" (Searle, 2004b: 83).[6] Este realismo mentalista puede remontarse a Searle (1984b), donde se nos dice que es un hecho puro y simple –"just a plain fact about the world" (Searle, 1984b: 15 del original, 20 de la traducción)– que el mundo contiene estados mentales conscientes. Dicho realismo mentalista es asimismo un realismo ingenuo: Searle (1984a) defiende que los estados mentales consientes tienen de hecho las propiedades que parecen tener, dado que por lo que a ellos respecta no hay diferencia entre cómo son las cosas y cómo nos parecen, motivo por el cual ha dicho en repetidas ocasiones (v. g. Searle, 1989a: 203; 1992: 122 del original, 131 de la traducción; 1997a: 112 del original, 118 de la traducción) que en el caso de la conciencia la única realidad es la apariencia. En resumidas cuentas, y extractando lo hasta aquí señalado, Searle trata de eludir –y, en la medida de lo posible, también de elidir– tanto posturas escépticas o relativistas (en epistemología) como objetivistas o

[5] A este respecto cabe añadir que Searle, de acuerdo con su realismo, como apuntábamos, entiende que el mundo es cognoscible y, además, en una vena ilustrada y cientista, que ese conocimiento –si es que podemos apostar por él– será, principal si no –*casi*… Searle titubea en este punto– exclusivamente, científico. Vid., v. g., Searle (1999b), un artículo en el que Russell parece hablar por boca de Searle para decir cosas tales como que "Tan pronto como logramos una forma sistemática de responder a una pregunta y conseguimos una respuesta que todos los investigadores competentes en ese área acuerdan que es la respuesta correcta, dejamos de considerarla una pregunta filosófica para considerarla científica" (Searle, 1999b: 2) –en la misma línea, vid. los primeros compases de Sarle, 1998, v. g.: "Science is systematic knowledge; philosophy is in part an attempt to get us to the point where we can have systematic knowledge (…); as soon as we think we really know something we stop calling it philosophy and start calling it science" (Searle, 1998: 380).

[6] Las cursivas son de Searle.

positivistas (en ontología). Dicho lo dicho, a nadie puede extrañarle la siguiente afirmación: Searle divide el mundo en dos compartimentos, y en cada uno de ellos mete un tipo de equipaje que –y esto le resulta obvio– es esencialmente diverso dado el carácter subjetivo de la experiencia consciente y la inaccesibilidad directa a la misma desde el punto de vista de la tercera persona. Searle traza así un esquema según el cual en el mundo existen fenómenos dependientes del observador y fenómenos independientes del mismo (para esta distinción, vid. Searle, 1995a).[7] Además, entre los segundos encontramos fenómenos ontológicamente objetivos y fenómenos ontológicamente subjetivos. A esclarecer la naturaleza de éstos y dilucidar el modo adecuado de aproximarnos a ellos ha dedicado Searle gran parte de su trabajo en filosofía de la mente. Los próximos apartados tienen por objeto la exposición y crítica de ese esclarecimiento y esa dilucidación.

### 1.1._El naturalismo de Searle

Searle afirma que su intención es la de ubicar la conciencia dentro de nuestra visión científica contemporánea del mundo. Según él, al menos dos de los componentes de esta visión del mundo están tan bien establecidos y son tan fundamentales que no resultan opcionales para los ciudadanos educados de nuestro tiempo: se trata de la teoría atómica de la materia y la teoría evolucionista en biología, dos elementos de la señalada visión científica del mundo que, entiende, hemos de tener en cuenta de cara a situar la conciencia dentro de nuestra concepción del mundo (Searle, 1992: 86 del original, 98 de la traducción).[8]

Respecto de la teoría atómica de la materia –el primero de los señalados componentes–, apuntemos brevemente que Searle no integra en su planteamiento del problema de la conciencia ningún elemento técnico concreto tomado de la física, sino que, desde una prudente distancia, traza una suerte de *grueso contorno causal* para su "modelo": hablará de niveles de explicación en física, los cuales tendrán en el planteamiento de Searle más que ver con la descripción que hacemos de las causas que con la cantidad o los tipos de éstas. En este sentido, distinguirá entre explicaciones causales macro-macro

[7] Se trata, evidentemente, de una distinción antes epistemológica que ontológica, lo cual no implica que no se halle ella íntimamente relacionada con la ontología de lo mental que Searle defenderá.
[8] En Searle (2005) las neurociencias aparecerán como el tercer –y decisivo– eje en torno al cual debemos hacer orbitar nuestra visión del mundo –y, asimismo, como el referente fundamental de su concepción de la conciencia y el problema mente-cuerpo: los "hechos básicos", según este texto, han de ser enfocados desde la base ofrecida por la física atómica, la biología evolucionista y las neurociencias.

(v. g., "el agua hirvió porque la calentaste") y explicaciones causales micro-macro (v. g., "el agua hirvió porque los rápidos movimientos de sus moléculas hicieron que la presión interna del líquido alcanzara el umbral de la presión externa del aire"), y sumará a esta distinción una perspectiva temporal según la cual las causaciones micro-macro acaecen sin diferencia temporal, es decir, sin claros antecedentes y consecuentes –tipo *bola de billar*–, sin que *explanans* y *explanandum* se hallen ubicados en cortes temporales –palmariamente– distintos (v. g. "la liquidez del agua es causada por la conducta y las propiedades de las partículas que la componen o integran"). Como parece poder apreciarse atendiendo a los dos primeros ejemplos, las causas no son diferentes: es nuestra descripción la que varía. A lo apuntado añade Searle que en el caso de explicaciones causales micro-macro necesitamos hacer referencia a complejas teorías que dan cuenta del modo en que el comportamiento de elementos del micronivel causa los rasgos superficiales que observamos en el macronivel –este sucinto esquema, como veremos, es de gran relevancia dentro del planteamiento del problema de la conciencia elaborado por Searle.

Por lo que toca al segundo de los señalados componentes de nuestra visión científica del mundo –la teoría de la evolución–, Searle destaca su importancia en relación con el problema de la conciencia de un modo muy conciso: determinados animales poseen determinados tipos de estructuras nerviosas capaces de causar y mantener procesos y estados conscientes, y tanto aquellos animales como estas estructuras surgieron evolutivamente. Así, propone, la conciencia y sus diferentes características deben ser comprendidas como fenómenos tan biológicos –y por tanto físicos (Searle, 1992[9]; 1995a; 1997a; 1998[10]; 2002b[11]; 2004a)– como la digestión o la fotosíntesis.

Searle se defiende –de antemano– de posibles críticas relativistas a la fundamentación cientista de su filosofía de la mente, dado que, entiende, podría argüirse que dicha fundamentación juega en su contra por cuanto la refutación de las teorías científicas en las que basa sus propuestas filosóficas supondría la invalidación de éstas. Su defensa consiste en reiterar su posición antirrelativista defendiendo la existencia de *hechos básicos* y presentando al relativismo como una postura incongruente: puede, argumenta, que nuestro conocimiento científico del mundo no sea más que un constructo temporal, fali-

[9] "[C]onsciousness is just an ordinary biological, that is, *physical*, feature of the brain" (Searle, 1992: 13 del original, 27 de la traducción; las cursivas son de Searle).
[10] "Consciousness is an ordinary biological, and therefore physical, feature of the organism, as much as digestion or photosynthesis" (Searle, 1998: 384).
[11] "[C]onsciousness is a mental and therefore biological and therefore physical feature of the brain" (Searle, 2002b: 61).

ble y, sin duda, perfectible, pero el hecho de que determinados segmentos del mismo sean reelaborados es un argumento antirrelativista y no al contrario, pues sólo bajo la suposición de la existencia de un mundo y una realidad absolutos y no-relativos tiene sentido semejante tarea de reelaboración. "El hecho de que cambien las opiniones es un argumento en contra del relativismo, no a su favor. (…) Los hechos no cambian, pero el alcance de nuestro conocimiento sí" (Searle, 2005: 333τ). No obstante, cabría presentar objeciones similares a la fundamentación cientista de Searle sin atisbo alguno de sesgos relativistas, pero no seremos nosotros quienes le ataquemos por ahí, pues, por una parte, su compromiso con las disciplinas y teorías que son el caso no presenta un cariz muy osado, dado que se adscribe a marcos muy generales y en ningún caso a tesis concretas, y, por otra –la principal–, la evidencia en favor de las mismas es, como Searle subraya, abrumadora: no se atalayan visos de que quepa la posibilidad de que pueda sucederle con ellas algo remotamente parecido a lo que le sucediera a Kant con la física de Newton y la lógica de Aristóteles –recuerde el interesado en atacar a Searle por este flanco la laxitud con la que éste pinta el cuadro de la física y la biología al que se adhiere: ¡se trataría de argumentar en una línea que abriera las puertas a dudas razonables sobre la teoría de la evolución y la naturaleza atómica de la materia!

## 2._La conciencia según Searle

Comencemos por el término «conciencia». Searle no ofrece una definición precisa del mismo, y esto no le parece un problema. Tampoco a nosotros nos lo parece: nadie ha ofrecido hasta la fecha una definición a la que resulte dable aspirar a candidata al unánime consenso y, como sugiriéramos ya en el capítulo segundo de la primera parte, no parece que este hecho vaya a suponer óbice en nuestro camino hacia una comprensión cabal del fenómeno de la conciencia. En esta línea, tampoco Searle parece preocupado al argumentar que la ausencia de una definición precisa no ha de suponer un impedimento ineludible para el estudio científico de la conciencia.[12] No obstante, su planteamiento de este extremo depende de una distinción que, a pesar de su carácter intuitivo, hace uso de nociones problemáticas (como las de esencia o sentido común): según la misma, existen, por una parte, definiciones analíticas, cuyo objetivo es analizar y hacer

[12] "Like most words, 'consciousness' does not admit of a definition in terms of genus and differentia or necessary and sufficient conditions" (Searle, 1993a: 7). Searle alude aquí, además de a las condiciones necesarias y suficientes, nociones ubicuas que atraviesan la historia de la filosofía y la de la lógica, a las condiciones aristotélicas de la definición: género y diferencia (vid. *Metafísica*, VII, 12, 1028 a – 1041 b).

explícita la esencia de un fenómeno y, por otra, definiciones de sentido común, las cuales simplemente identifican el objeto al que aluden y aclaran de qué hablamos al referirnos a él.[13] La definición de «conciencia» caería en la propuesta de Searle del lado de las segundas. Así, subraya que no debemos esperar de la definición de conciencia con la que pretendamos trabajar una precisión asimilable a la de las definiciones que típicamente aparecen al final de una investigación científica, sino que habremos de conformarnos con una definición –de sentido común– suficiente para identificar el objetivo de la investigación (Searle, 1998: 381; 2000a: 38), dado que Searle entiende que este tipo de definición ofrece exactamente lo que necesitamos en los primeros pasos de la investigación de cara a avanzar hacia una explicación científica de la conciencia. Los términos que formen parte de tal definición tendrán pues el significado propio del lenguaje ordinario antes que el propio de algún lenguaje técnico específico. De este modo, «conciencia» se refiere, según Searle, a aquellos estados del sentir y del advertir –estados de sensación y percatación o alerta (*awareness*)– que comienzan cuando despertamos de un sueño sin ensueños y se prolongan hasta que volvemos a dormirnos, caemos en coma o morimos (Searle, 1990b: 635; 1992: 83 del original, 95 de la traducción; 1993a: 7; 1997a: 5 del original, 19 de la traducción; 1998: 381; 2007a: 326; 2007b: 170).[14] Es interesante hacer notar que la definición de conciencia que Searle ofrece abre el paso tanto a estados de actitud proposicional (como deseos o creencias) como a estados de ánimo, percepciones o sensaciones corporales –esto le llevará en el tratamiento del problema de los *qualia* a optar por una concepción amplia de la extensión de este término que, acepta pero considera potencialmente engañoso–. El acento lo pone Searle, en cualquier caso, en el carácter subjetivo y cualitativo de la experiencia consciente. La conciencia nos es así presentada por el de Denver como un fenómeno interno esencialmente cualitativo y subjetivo, un fenómeno con una ontología de primera persona – Searle, por otra parte, ha hecho hincapié en la diferencia entre la conciencia así concebida y la conciencia moral, la atención, el conocimiento (Searle, 1993a) o la auto-

[13] Searle introduce está distinción, por ejemplo, en Searle (1998: 381; 2000a: 31).

[14] En Searle (2000a) el filósofo de Denver ha matizado su posición en este extremo indicando que el término «conciencia» se refiere a estados y procesos internos, subjetivos y cualitativos de sensación, percepción, pensamiento y alerta (awareness). Su definición, en muchas ocasiones repetida, es así, en su última formulación –que incluye, a diferencia de versiones anteriores, la noción de "feeling"–, la siguiente: "*Conscious states are those states of awareness, sentience or feeling that begin in the morning when we wake from a dreamless sleep and continue throughout the day until we fall asleep or otherwise become "unconscious"*" (Searle, 2007a: 326 –cursivas en el original). En Searle (2007b) la definición es prácticamente identica: "Consciousness consists of those subjective states of sentience or feeling or awareness that begin when we wake from a dreamless sleep and continue on throughout the day until we become unconscious again" (Searle, 2007b: 170).

conciencia[15] (Searle, 1993a: 8; 1997a: 6 del original, 20 de la traducción) y, asimismo, ha distinguido alerta (*awareness*) de conciencia (*consciousness*), vinculando aquélla más bien con aspectos cognitivos de la vida mental (Searle, 1992: 84 del original, 96 de la traducción) y, no obstante, conceptualizándola como parte de su definición de conciencia: «awareness» caería dentro de la noción general de «consciousness» (estados del sentir y del *advertir*, recuérdese), mientras no sucedería lo mismo a la inversa.

De forma más explícita, Searle ofrece una caracterización de la conciencia en la que los siguientes rasgos juegan un papel fundamental: carácter cualitativo, subjetividad y unidad. Esta triada de rasgos (*subjetividad cualitativa unificada*) constituye lo que Searle ha denominado "la esencia de la conciencia" (Searle, 2000a: 42τ; 2007b: 170τ), aquello que hace de ella lo que es al otorgarle su carácter distintivo. Searle (2000a; 2004a; 2007b) enfatiza la imposibilidad de disociar tales rasgos: no se trataría siquiera de características diferentes o separadas de la conciencia, sino de distintos aspectos de uno y el mismo fenómeno.[16]

En cuanto a los rasgos secundarios que, junto con la referida triada de rasgos primarios, integran lo que Searle denomina *la estructura de la conciencia* (Searle, 1992), admite el de Denver no disponer –en su primera elaboración sistemática de la cuestión– de una concepción acabada de al menos dos de ellos: se trata de la temporalidad –acerca de la cual ofrecerá algunas pistas en Searle (2004a)– y la sociedad –de la que tratará en Searle (1995a)–. Por lo que al resto toca, cabe hacer notar que los planteamientos de Searle en relación con los mismos, a pesar de ocupar un segundo plano, no dejan de incluir, como enseguida veremos, algunas apuestas descriptivas de considerable relevancia.

Comenzaremos –por motivos de conveniencia expositiva (*ordo naturalis* sólo aparente, pues)– a presentar su caracterización de la conciencia partiendo del primero de los rasgos citados: el carácter cualitativo de la conciencia. Respecto del mismo, Searle destaca que la conciencia se presenta en un número de modalidades limitadas o finitas: v. g.,

[15] Noción que Searle (vid., v. g., Searle, 2007b: 170) entiende como un elemento posible –y suficiente– pero no necesario de la conciencia (toda autoconciencia es consciente pero no a la inversa) y que, ocasionalmente, ha dotado de un sentido estrecho (*narrow*) caracterizándola como un estado mental consciente en el que la atención se dirige no a los objetos de la experiencia sino a la experiencia consciente misma (Searle, 1992: 143 del original, 152 de la traducción).

[16] De este modo pretende Searle minimizar las posibilidades de ser malinterpretado: subraya (Searle, 2004a: 136; 2007b: 170) que tal vez no hubiera acertado en sus anteriores aproximaciones (vid., especialmente, Searle, 1992) a detectar y presentar adecuadamente el carácter unitario de estos tres rasgos fundamentales.

sensitivas, intelectivas o emocionales.[17] Hay que apuntar que Searle abandonaría esta nomenclatura de *modalidades finitas* (Searle, 1992), sustituyéndola –a partir de Searle (2000a), artículo aparecido en *Annual Review of Neurosciences* y recogido posteriormente en Searle (2002a)– por la más general y explícita de *carácter cualitativo*. Este carácter cualitativo de la conciencia tiene que ver, en efecto, con la noción de *qualia*. Searle definirá el referido carácter cualitativo en continuidad con el clásico *what it is like* de Nagel (Nagel, 1974): "Para cada experiencia consciente hay algo que se siente como, o algo que es como, tener dicha experiencia consciente" (Serale, 2000a: 39τ). Respecto de la noción de *qualia*, aprovechemos para incidir brevemente en la amplitud de la que Searle ha dotado al término: a diferencia de la mayoría de los *qualófilos*[18] –excepción habida, por ejemplo, de Strawson (1994) o Flanagan (1992)– Searle propone que no sólo las experiencias perceptivas o los estados anímicos están dotados de aspecto cualitativo, sino que todo el campo de la conciencia lo está, pues toda experiencia consciente es siempre cualitativa, incluyendo tanto los segmentos perceptivos y emocionales como los cognitivos o intelectivos. En este sentido, Searle considera que hay, como apuntábamos, cierto potencial engañoso en la noción de *qualia*, dado que –teniendo en cuenta el modo en que habitualmente se utiliza en la filosofía de la mente contemporánea– puede llevar a pensar en la existencia de estados conscientes no cualitativos (Searle, 2004a: 134; 2007b: 170), mientras, según Searle, la forma correcta de tratar el asunto sería mostrar el modo en que conciencia y *qualia* son de hecho términos coextensivos (Searle, 2007a: 326). ¿Y en qué sentido es cualitativo un pensamiento?, cabría preguntar en vista de que, intuitivamente (o, más bien, según parecen haber intuido gran cantidad de filósofos de la mente contemporáneos), ver el cielo azul o estrellado, oler nuestra comida favorita o saborearla o, simplemente, sentirnos tristes, cuentan como estados mentales portadores de un verdadero carácter cualitativo mientras que pensar en un teorema no parece formar parte de esa clase de estados mentales que experimentamos como dotados de ese algo que nos lleva a decir que hay algo que es como sentirlos o estar en ellos. Searle despacha la cuestión con un ejemplo –que cabe escoliar como sigue: pensar la misma operación matemática en diferentes idiomas es un acto cualitativamen-

[17] En cuanto a las primeras, incluirían las modalidades de los cinco sentidos, además de la propiocepción, la interocepción o el sentido del equilibrio. Las segundas estarían en la propuesta del de Denver vinculadas con su noción de *flujo del pensamiento*, del que formarían parte *palabras* e *imágenes* (no sólo visuales, sino también auditivas e incluso olfativas). Las terceras, que Searle no presenta como disociadas o separadas de las anteriores, aludirían a todo tipo de sentimientos y estados anímicos, desde el miedo o la tristeza a la ira o la rabia.

[18] Término cuya acuñación atribuye Levine (1997; vid. Levine, 1994 para una aplicación del término) a Dennett.

te diverso– del que pretende servirse para ilustrar que todo pensamiento, al igual que toda experiencia consciente, tiene, como una propiedad ineludible, una determinada perspectiva –consecuencia, en el planteamiento de Searle, del hecho de que toda intencionalidad se presenta con un determinado aspecto (Searle, 1992: 131 del original, 140 de la traducción) y un determinado carácter cualitativo–. Indiquemos para acabar que según Searle (2000a) este rasgo diferencia a la conciencia de cualquier otra característica del mundo: que un organismo sea consciente significa que hay algo que es como ser él, cosa que no puede decirse de ninguna entidad no-consciente.

En relación con el segundo rasgo, la subjetividad de la experiencia consciente, lo primero que cabe apuntar es que Searle ha reservado para este aspecto de la conciencia un lugar privilegiado en su concepción de la misma. No en vano gran parte de sus esfuerzos teóricos están destinados a presentar a la subjetividad como un componente real del mundo real –valga el pleonasmo–, y de ahí que hable de "*subjetividad ontológica*". La subjetividad, como indicábamos, ocupa un lugar privilegiado en la concepción de la conciencia que Searle ha venido desarrollando, y es que, desde su punto de vista, lo definitorio de un estado consciente, lo que confiere a un estado consciente su identidad en tanto tal, es el hecho de que tal estado sea sólo introspectivamente cognoscible y requiera por tanto de la presencia de un sujeto capaz de sentirlo o experimentarlo: se hace imprescindible, pues, la presencia de un sujeto, lo cual hace explícito el sentido en que la subjetividad define a la conciencia. Este rasgo, por otra parte, se halla en el planteamiento de Searle inextricablemente vinculado con el anterior: aquél implica a éste, pues toda experiencia cualitativa es –necesariamente– experimentada en primera persona por un sujeto. Searle traza pues un vínculo entre estos dos primeros rasgos al hablar del *sentimiento subjetivo* de la conciencia, indicando que la subjetividad de la conciencia involucra la cuestión del *what it is like* y añadiendo que es este rasgo de la subjetividad el principal responsable del problema filosófico de la conciencia. La conciencia, en definitiva, existe según Searle como subjetividad: su modo de existencia es subjetivo. En Searle, la locución "subjetividad ontológica" pretende significar que la realidad incluye determinados estados de cosas que sólo pueden existir en cuanto experimentados por determinados entes (*subjectum*) para los cuales hay modos determinados en que se sienten o experimentan tales estados (*qualitas*), motivo por el cual hablará Searle de una *ontología de primera persona*: la ontología de los estados experimentados por sujetos que tienen como característica distintiva la posibilidad de tener experiencias conscientes –experiencias que son suyas y existen *para* ellos: ningún otro puede experimentarlas; la subjetividad *existe*

*para* el sujeto–. Que la experiencia sea esencialmente subjetiva quiere decir aquí que sólo *yo* experimento mis experiencias. Se trata, en resumidas cuentas, de una aproximación ontológica a la subjetividad que incide en la peculiaridad de la misma en tanto forma de existencia antes que en tanto forma de conocimiento. Epistémicamente subjetivo se referiría, por contraposición a ontológicamente subjetivo, al estatus de conocimientos o proposiciones dependientes de actitudes, sesgos o sentimientos de los sujetos. Partiendo de esta distinción, Searle cree hallarse en disposición de afirmar que "la subjetividad ontológica de un objeto de estudio no excluye una ciencia epistémicamente objetiva de ese mismo objeto de estudio" (Searle, 2004a: 136τ).

El tercer rasgo de la caracterización de Searle, la unidad de la conciencia, se refiere al hecho de que los diversos aspectos perceptivos, intelectivos, atencionales y emocionales de una y la misma experiencia consciente se nos dan, por así decir, juntos: como formando parte de "uno y el mismo evento consciente" (Searle, 1992: 130 del original, 139 de la traducción). Searle hablará de unidad vertical y unidad horizontal. La última tiene que ver con la memoria icónica y la memoria a corto plazo y se encuentra relacionada con el mantenimiento y la organización integrada de experiencias conscientes en periodos cortos de tiempo (Searle usa como ejemplo una frase durante el tiempo en que es pronunciada), mientras la primera lo está con la unificación en la experiencia consciente de diversos aspectos de la misma (v. g., el frío que siento mientras camino alegre por la calle pensando en el mejor camino para llegar al lugar en el que aparqué). Por otra parte, según Searle (2000a), la unidad de la experiencia consciente viene ya dada o está ya implícita en el carácter cualitativo y subjetivo de la misma: no cabe imaginar un sujeto cualitativa y subjetivamente consciente cuyo estado consciente se encuentre fragmentado, porque, según Searle, tal situación imaginaria estaría presentándonos diferentes conciencias con sus peculiares puntos de vista, esto es, diferentes centros de conciencia, de donde extrae la conclusión de que no existe la posibilidad de obtener carácter cualitativo y subjetividad sino dentro de los límites de una tal unidad.

Searle, como indicábamos, hace uso en su propuesta descriptiva de una serie de rasgos secundarios que vienen a sumarse a la triada de la que acabamos de ocuparnos para conformar lo que denomina *la estructura de la conciencia*. Nos hemos referido a ellos como secundarios queriendo con ello resaltar que no integran –directamente– lo que Searle, como hemos visto, denomina *la esencia de la conciencia*; pero que sean en este sentido secundarios no implica que carezcan de relevancia dentro de la propuesta

descriptiva de Searle. Muy al contrario: estos rasgos secundarios le sirven a Searle para introducir alguna de las claves de su concepción de la conciencia, y por ello nos detenemos en este punto a presentar una lista numerada y comentada de los mismos.

1._ *Intencionalidad*. Se trata de una de las nociones fundamentales de la filosofía de la mente contemporánea y ocupa un lugar central en la concepción de la conciencia de Searle. Su definición de intencionalidad como "propiedad de muchos estados y eventos mentales por la cual éstos están dirigidos a, o son sobre o de, objetos y estados de cosas en el mundo" (Serale, 1983: 1 del original, 17 de la traducción) es lo suficientemente general como para producir más bien pocas disensiones. Sí cabría, por el contrario, esperar que las produjeran afirmaciones como las contenidas ya en el prefacio a la obra que acabamos de citar, afirmaciones que presentan a la intencionalidad de lo mental como "originaria", "genuina", "auténtica", "básica", "original" y previa o anterior a cualquier otra clase de intencionalidad y que desembocarán en un credo según el cual lo único "esencialmente intencional" son los estados mentales conscientes. Podríamos, en cualquier caso, glosar la definición searleana de intencionalidad indicando –en términos que trazarían una línea que se prolonga de Brentano a Husserl–[19] que toda conciencia es conciencia-de. Pero no se trataría de una glosa enteramente justa, porque Searle –abanderado en este punto de la línea canónica en filosofía de la mente–[20] contempla excepciones a esta regla, como los sentimientos o estados de ánimo del tipo de la angustia inmotivada o la depresión melancólica, en los cuales se está ansioso o deprimido por nada en particular, de forma que no se puede definir con claridad el objeto de tales estados mentales. Con todo, Searle sostiene que mientras todo estado consciente es en general intencional, son muchos los ejemplos que podrían darse de estados intencionales inconscientes –aunque todos los que Searle, finalmente, trae a colación son ejemplos de actitudes proposicionales no actualmente conscientes: v. g., podría decirse que yo deseo obrar bien o que creo que Orión es a la vez un gigante mítico nacido de los orines de Zeus, Hermes y Poseidón y una constelación cuyas estrellas principales están a una distancia de la Tierra de entre 800 y 1500 años luz, aunque ese deseo y esa creencia no pululen actualmente por mi mente consciente.

[19] Vid., v. g., Albertazzi (1996: 199).
[20] Kenny (1963) podría considerarse como inaugurador de la tendencia heterodoxa en este extremo al analizar las emociones como sentimientos intencionales (intencional feelings).

Un aspecto relevante del uso que Searle hace de la noción de intencionalidad en su descripción de la *estructura de la conciencia* es que, en su planteamiento, intencionalidad y conciencia aparecen ligadas por un vínculo necesario:[21] toda intencionalidad es aspectual –todo estado intencional tiene, según Searle, un cierto contorno de aspecto– y toda experiencia consciente está así dotada de perspectiva –en la conciencia, los objetos intencionales no se presentan sino bajo aspectos: el mismo objeto puede ser deseado, recordado, percibido, etc., y cada una de esas modalidades intencionales o psicológicas interviene de forma peculiar en la experiencia consciente; además, incluso dentro de una misma modalidad (v. g., desear), uno y el mismo objeto puede ser representado y experimentado en diferentes estados mentales bajo diferentes aspectos (v. g., desear una naranja no equivale a desear el fruto del *citrus sinensis*: varía, por así decir, el modo de presentación; varía, pues, algo análogo al *Sinn* fregeano (Frege, 1892), y aunque el objeto al que ambos estados intencionales apunten sea el mismo, en cada caso el contenido es distinto)–. Este vínculo entre conciencia fenoménica e intencionalidad se hace explícito en la distinción que Searle traza entre intencionalidad intrínseca u original, intencionalidad derivada e intencionalidad "como-si". Tal y como Searle presenta la distinción (Searle, 1992, 1997c, 2004a) las cosas estarían del siguiente modo: sólo la mente *tiene* verdadera intencionalidad (*original*). Por su parte, el tipo de intencionalidad que signos y símbolos poseerían –y en particular la intencionalidad que presenta el lenguaje– deriva de aquélla, mientras que el tercer tipo responde a atribuciones metafóricas (Searle, 1980: 188 ya apunta a este tipo de atribuciones), como, por ejemplo, cuando "prosopopeyizamos" o "psicologizamos" artilugios: «parece que mi ordenador *quiere* tomarse un descanso» sería un ejemplo de atribución intencional "como-si".

Por último, Searle (1983, 1992, 2004a) defiende que los estados intencionales no pueden ser cabalmente considerados de forma aislada: no podrían ser los estados intencionales que son de nos ser por su relación con otros estados intencionales. Así, una creencia es el estado intencional que es dada su relación con otros estados intencionales, es decir, dada su posición en una *Red* de estados intencionales. Por otra parte, los estados intencionales, según Searle, necesitan para contar como tales –y como el estado intencional concreto que cada uno de ellos pueda ser– no sólo de la referida interrela-

[21] Al punto que, en la propuesta de Searle, conciencia e intencionalidad parecen apuntar una a la otra y requerirse mutuamente en tal medida que considera el de Denver que una teoría completa de la intencionalidad exige una explicación de la conciencia, o, en la misma línea –pero virando de las implicaciones epistemológicas de su tesis a las ontológicas–, que sólo un organismo que pueda tener estados intencionales conscientes puede tener estados intencionales. Este vínculo, como veíamos en el capítulo octavo, es el que Searle intenta establecer mediante su Principio de Conexión.

ción, sino además de una base de capacidades y habilidades no intencionales que denomina *Trasfondo*.[22]

2._ *Desbordamiento*. En vaga conexión con la apuntada noción de *Red* hallaríamos el modo en que Searle presenta este rasgo invitándonos a pensar en el sentido en que explicitar el contenido de cualquier estado mental consciente resultaría, en cualquier caso, una tarea poco menos que eviterna, dado que el *contenido inmediato* tiende a rebosar, conectando con otros contenidos en cadenas de longitudes y ramificaciones imprevisibles.

3._ *Familiaridad*. Con este rasgo Searle pretende subrayar que nuestras experiencias conscientes se dan, por así decir, dentro de unos márgenes que nos resultan familiares. Un caso extremo puede servir para ilustrar el punto al que Searle quiere conducir nuestra atención: incluso un adulto que jamás haya visto el mar tendrá al verlo por primera vez una experiencia que se desarrollará dentro de un *aspecto de familiaridad* –será una experiencia visual, enmarcada espacialmente con arreglo a las leyes de la óptica, la perspectiva, la gravedad, etc., y aunque puede que las imágenes o los olores le resulten nuevos y sorprendentes, no dejarán de ser imágenes y olores, fenómenos concretos (sorprendentes o no, nuevos o no, pero, en cualquier caso, reconocibles) dentro de modalidades perceptivas que el sujeto conoce–, el cual no concibe Searle como un dato, sentimiento o rasgo separado o separable de la conciencia: no siento por una parte dolor en el brazo y por otra *familiaridad de dolor* y *familiaridad corporal*. Además, más allá de lo que nuestro ejemplo sugiere, la familiaridad de nuestras experiencias conscientes tiene que ver con la organización general de las mismas, la cual incluiría la familiaridad que tenemos con el famoso *what it is like*, con eso que es como ser yo, con el modo particular en que siento mis experiencias y me siento a mí mismo (Searle, 1992: 134 del original, 143 de la traducción).

4._ *Centro y periferia*. Searle traza esta distinción atendiendo a conceptos que no resulta difícil explicitar: la atención tiene límites y en su núcleo "caben" una cantidad discreta de "cosas", pero sus márgenes no son enteramente inconscientes, pues en ellos encontramos estados o partes de estados mentales que resulta comprometido desterrar, concebir como ajenos o externos al estado actual de conciencia del sujeto que sea el

[22] Consideramos innecesario, dado el objeto de nuestro trabajo, inventariar el aparato descriptivo desarrollado por Searle para el análisis de la intencionalidad (vid., especialmente, Searle, 1983 –algún añadido puede hallare en Searle, 2004a), dentro del cual destacan las nociones de *dirección de ajuste* y *condiciones de satisfacción* –esta última, anotemos al margen, está estrechamente relacionada con las de *Red* y *Trasfondo*, a las que acabamos de aludir y de las que nos ocuparemos más adelante.

caso. El objetivo de Searle en este punto es el de matizar la contraposición consciente/inconsciente mediante la contraposición entre los contenidos que ocupan el centro de un estado atencional y los que constituyen su periferia, en ocasiones en forma de contexto sólo *presente* de la forma más sutil.

5._ *Gestalt*. En pocas palabras, cuando Searle –usando esta clásica noción psicológica– hace referencia a la *estructura gestáltica de la conciencia*, quiere significar que las percepciones y, en general, los estados mentales, se dan organizados: así, a pesar de que mi sistema nervioso "procese" por separado la información visual de forma, color y movimiento, yo veo a un perro marrón caminando hacia el extremo de la calle, es decir, un objeto definido y organizado como un todo en una escena definida e igualmente organizada. Además, y en línea con lo señalado en el punto anterior, Searle incluye en este rasgo –en continuidad con las categorías gestálticas de fondo y figura– el hecho de que, tal y como propone, para cada estado mental consciente existe un centro –figura– que resalta sobre un fondo (Ibíd.: 133 del original, 142 de la traducción).

6._ *Condiciones límite*. El significado de este rasgo de la caracterización searleana de la *estructura de la conciencia* puede expresarse en términos muy generales como sigue: toda experiencia consciente no patológica tiene –generalmente a modo de telón de fondo no tematizado: inadvertido ruido de arrastre, armonía de las esferas, para expresarlo líricamente– un carácter situado o localizado. Searle propone como una suerte de *condición límite* de la conciencia este carácter: la localización espaciotemporal y sociobiológica de cualquier estado mental consciente no necesita hallarse en el centro ni en la periferia de un estado mental consciente para ejercer su efecto –v. g., "sé" que soy Asier Arias, cordado, mamífero, *Homo sapiens*, ciudadano de un estado europeo del siglo XXI, también "sé" qué día de la semana es hoy, y que ya hace bastante tiempo que almorcé, pero no el suficiente como para comenzar a preparar la cena, y este tipo de conocimiento límite, que lo invade todo sin dejarse ver por ninguna parte, sólo negativamente se muestra con efectiva patencia, es decir, sólo cuando el mismo falla o falta: en estados de desorientación.[23]

7._ *Estados anímicos*. Al tratar de ofrecer sus perspectivas acerca de la relación entre conciencia y estados anímicos Searle viene a desembocar en un lugar similar a

[23] Cabe añadir que Searle (1994a: 70) apunta a las diferencias habidas entre la forma de aprehender la situación –dentro de los márgenes de estas condiciones límite– entre animales no humanos y humanos como una de las claves que diferencian la mente y la conciencia humana de las del resto de los animales –no obstante, Searle pone el acento en este trazado de fronteras en la peculiaridad y las potencialidades del lenguaje.

aquél al que le condujera su reflexión acerca del carácter cualitativo de la conciencia. En ambos casos, desde su punto de vista, la pregunta a formular sería la siguiente: ¿existe algún estado mental consciente no cualitativo (*qualia*) y/o sin un *sabor* o *tono* (Searle, 2004a: 139) afectivo (estados anímicos)? En cualquiera de los casos, su respuesta es no. Desde su punto de vista, incluso cuando no alcanzamos a percibir con distinción una inclinación concreta de nuestro ánimo (hacia la alegría o la tristeza, por ejemplo), incluso cuando nos mostramos incapaces de discernir en nosotros un estado anímico evidente –ocasiones en las que diríamos algo como: "no, no sé… no estoy ni especialmente contento ni decididamente triste, pero tampoco indiferente"– ese tono está presente. Por otra parte, Searle (1992), como el grueso de los filósofos de la mente contemporáneos, trata a los estados anímicos como estados mentales –bien cabe que resulte más correcto en este punto hablar de aspectos de estados mentales– no directamente intencionales o no intencionales en sí mismos.

8._ *Dimensión de placer/displacer*. Este rasgo de la caracterización de Searle guarda, obviamente, una estrecha relación con el anterior. Pero no debemos precipitarnos, pues Searle los trata por separado y, de hecho, afirma (Searle 2000a: 45) que la pregunta acerca de los estados anímicos difiere de la pregunta acerca de esta dimensión. Así, desde su punto de vista, la pregunta acerca de lo agradable o desagradable de una experiencia consciente cualquiera difiere de la pregunta acerca del estado anímico que la "acompaña", aunque la relación entre ambos horizontes acostumbre a ser concordante. Además, propone, toda experiencia consciente transcurre dentro de unos márgenes en algún sentido impuestos por este rasgo: siempre existe esa dimensión y, por tanto, siempre podemos situar cualquier estado consciente, concebido holísticamente, en un continuo que iría del polo agradable a su opuesto desagradable. Todo estado mental consciente está, en su totalidad, embebido en esta dimensión de placer/displacer.

9._ *Percepción/voluntad*. Searle (2004a: 142) traza una distinción entre la experiencia consciente de percepciones y voliciones (el problema de las cuales trata en Searle, 2004b), entre la experiencia pasiva de la percepción y la experiencia de la actividad volitiva. A pesar de que admite que ambas categorías pueden interpenetrarse causando cierta confusión acerca de sus límites, define la primera como la sensación que el sujeto tiene de que algo le sucede, y la segunda como la sensación que el agente tiene de que hace algo.

10._*El yo*. Dentro del debate contemporáneo acerca de la conciencia uno de los temas más abarcadores y complejos es, junto con el anterior extremo de la conciencia

volitiva y la libertad, el del yo y la identidad personal –al punto que la habitual especialización académica comienza a tratar ambos debates con cierta independencia y a elaborar agendas preñadas de cuestiones entre las que no resulta sencillo tender puentes–. Searle entiende que la cuestión del yo es de capital importancia para el problema de la conciencia, pero admite que su tratamiento de la misma (Searle, 2004a: 280 y ss.) no es sino una aproximación tentativa a un escurridizo tema que atraviesa la historia toda del pensamiento occidental. No entraremos a exponer –y, por tanto, tampoco a comentar– el tratamiento que Searle hace de este tradicional problema. No obstante, nuestra sumaria presentación del aparato descriptivo desplegado por el de Denver restaría incompleta en caso de que omitiéramos incluso la más lacónica de las alusiones a este punto. La referencia clásica es aquí Hume. Como es sabido, el escocés arremetió implacablemente contra la concepción sustancialista del yo: no encontró nada simple y continuo entre sus *impresiones* e *ideas* a lo que cupiera dar tal nombre y concluyó que, no pudiendo hallar en sí mismo nada más allá de tales *percepciónes*, no resultaba razonable hablar de un *yo* estable y constante más allá de aquel famoso *sí-mismo*-haz-de-percepciones. No pudiendo derivar la idea del yo de ninguna *percepción*, consideró que sólo cabía entenderlo, precisamente, como un sistema de percepciones (vid. Hume, 1738: Libro I, Parte IV, Sección VI).[24] Searle admite que no hallamos tal idea de yo por ninguna parte, pero, a diferencia de Hume, sí que encuentra la necesidad de *postular* (Searle, 2004a: 295) algo por encima del consabido haz humeano, y con esta necesidad de apelar al yo como *instancia formal*[25] se topa el de Denver por cuanto entiende que necesitamos dar cuenta de la acción, la libertad, la responsabilidad, la racionalidad, etc. Searle utiliza la imagen del *punto de vista* para exponer su punto de vista, valga la redundancia: siempre vemos desde un punto de vista, pero no vemos el propio punto de vista, dado que, como el yo, es una suerte de condición, una instancia o principio formal que nos hace falta para organizar nuestra experiencia y dotarla de sentido. Searle, en cualquier caso, presenta sus notas acerca de este tema en clave de breve aproximación tentativa y sugiere que la experiencia consciente y la identidad personal –entendida como la sensación consciente de ser yo– se hallan vinculadas por el peculiar modo en que sentimos que hay algo diverso

[24] Cierto es que en el Libro II (particularmente en la Parte I, Sección IX) parece olvidar Hume las conclusiones alcanzadas en el Libro I, así como lo es que en el *Apéndice* matiza su postura, pero, evidentemente, ésta es harina de otro costal: aquí nos interesa la alusión, no la discusión –por la orientación que aquélla comporta sin necesidad de incursionar los dilatados excursos a que ésta obligaría.

[25] Se trata de una noción que Searle utiliza de un modo ciertamente equívoco, pues con ella da la sensación de querer eludir la sustancialidad del yo para volver a caer en ella –al hablar del *yo* en términos, por una parte, de *principio formal* y, por otra, en términos de *entidad* (consciente, racional).

respecto del resto de presentaciones en el modo en que nos somos dados a nosotros mismos en primera persona (Searle, 2004a: 283) en la experiencia subjetiva, la cual, mediante la memoria, se relaciona tanto con la señalada unidad –horizontal– de la experiencia consciente como con la impresión de la unidad o continuidad de la identidad personal en el tiempo.

Planteado ya –por más que compendiosamente– el aparato descriptivo positivo elaborado por Searle, pasemos a ocuparnos brevemente de su reverso negativo: el modo en que completa su caracterización de la conciencia presentando los *rasgos tradicionales* que no atribuirá a la misma y los motivos que le llevan a excluir dichos rasgos de su caracterización. Searle rechaza explícitamente las nociones tradicionales de *incorregibilidad* e *introspección*. Respecto de la primera, Searle opina que, ciertamente, en el ámbito de lo consciente no tiene cabida la distinción entre realidad y apariencia que sustenta la posibilidad del error en otras esferas (v. g., el juicio) porque en dicho ámbito *la realidad es la propia apariencia*; y, sin embargo, se hace aquí igualmente posible el error, entiende Searle, en virtud de malas descripciones de nuestros propios estados mentales dadas circunstancias del tipo de la falta de atención sobre los mismo, malinterpretaciones o autoengaños. Vemos pues que, a diferencia de otros filósofos que también rechazan la incorregibilidad (como Dennett), Searle está pensando a la hora de depreciar este rasgo tradicional (vid. Searle, 1992: 146 y ss. del original, 154 y ss. de la traducción) en ejemplos de estados mentales complejos (vinculados antes con la así llamada autoconciencia y la identidad personal que con la percepción, la atención o la introspección inmediata) y ampliamente extendidos en el tiempo en los que la reflexión jugaría un papel fundamental (Searle habla de complicadas relaciones de pareja o filiaciones políticas para ilustrar su propuesta). Por lo que a la introspección se refiere, lo que Searle rechaza no es la concepción ordinaria según la cual puedo, por así decir, pensar sobre mis propios estados mentales, sino la idea de que existe una facultad especial para el acceso a la propia mente. Lo que pone en duda, pues, es la concepción de la introspección como metáfora visual y espacial de la capacidad de acceso a los propios estados mentales conscientes. Esta metáfora, propone, funciona mal tanto considerada desde el punto de vista espacial como desde el punto de vista visual porque en los actos de entrar (*acceder*) o ver hay una clara distinción entre el acto y el objeto, cosa que no puede decirse que suceda cuando pensamos o reflexionamos sobre nuestros propios estados mentales: en el caso de la atención dirigida a los propios estados mentales no

hay distinción entre el acto de observar y la cosa observada, con lo cual el referido modelo o metáfora visual, que funciona bajo la presuposición de una distinción de ese tipo entre la cosa vista y la visión de ésta, no sería en este punto más que una mala guía de interpretación, ya que toda introspección de un estado mental consciente es ella misma, precisamente, dicho estado mental.

### 3._ Naturalismo y conciencia en la filosofía de la mente de Searle

Visto ya el marco filosófico en que se encuadra la producción de Searle sobre el problema de la conciencia, y visto ya, asimismo, el marco descriptivo de su propuesta, pasemos a ocuparnos del modo en que integra Searle ambos extremos en su intento de elaborar una aproximación teórica apropiada al problema de la conciencia. Analizaremos, pues, el modo en que Searle aplica los supuestos de carácter general sucintamente presentados en el primer apartado de este capítulo al fenómeno descripto en el segundo. La clave de dicha aplicación puede hallarse en lo que Searle ha denominado *naturalismo biológico*, etiqueta que utiliza para definir su filosofía de la mente y, particularmente, su concepción de la conciencia. Necesitaremos transitar diferentes aspectos del planteamiento de Searle para dar contenido a la referida etiqueta. Consideramos que nuestra exposición será más clara si comenzamos por delinear los contornos de las cuestiones más abarcadoras y dejamos que los aspectos de detalle vayan surgiendo conforme avanzamos en la misma.

El objetivo fundamental del naturalismo biológico es el de desterrar el halo de misterio en el que habitualmente se ha presentado envuelta a la conciencia. Este objetivo tiene que ver con el modo en que tradicionalmente ha sido tratado el llamado problema mente-cuerpo, y Searle se muestra explícito respecto del mismo y las perplejidades que causa. Así, le encontramos ya en las primeras frases de *The Rediscovery of the Mind* afirmando que el problema mente-cuerpo tiene una solución simple y disponible para cualquier *occidental enculturado*,[26] una solución que, dice, todos sabemos que es cierta. Dicha solución reza como sigue: los fenómenos mentales son causados por procesos neurofisiológicos en el cerebro y son a su vez rasgos del cerebro (Searle, 1992: 1

[26] Searle propone, pues, su tesis de que la solución al problema mente-cuerpo se halla disponible no como una afirmación demostrable dentro de una teoría, sino como una consecuencia lógica de una *Weltanschauung*, un sistema de creencias que explica, mejor o peor, el mundo (Benítez, Fernández & del Arco, 2001: 266).

del original, 15 de la traducción).[27] El primer impedimento que Searle encuentra a la hora de extraer las implicaciones pertinentes de esta tesis se halla en el vocabulario tradicionalmente empleado en la formulación del problema mente-cuerpo, y en este sentido incide el filósofo de Denver en la necesidad de abandonar dicho vocabulario de cara a abordar con solvencia el problema de la conciencia. Para Searle, pues, "el lenguaje tradicional (…) es parte de la irresolubilidad del problema mente-cuerpo" (García Valero, 2003: 55).[28] Así, cuando a la hora de plantear el señalado problema la tradición preguntaba cómo algo mental puede ser físico, la propia utilización de las nociones de "físico" y "mental" creaba las paradojas habituales en el tratamiento teórico de dicho problema. De este modo, Searle encuentra que se hace necesaria una reelaboración conceptual y un cambio terminológico que propicie una redefinición de "físico" y "mental" mediante la cual dichas nociones no se presenten ya en el debate con sus tradicionales significados antagónicos. Asimismo, entiende que las categorías que han servido para trazar las fronteras entre bandos enfrentados en torno al señalado problema, categorías como monismo, dualismo o materialismo, requieren de un tratamiento analítico análogo, y en este sentido propone que tanto monistas como dualistas se hallan en un error, dado que han adoptado un vocabulario –que considera obsoleto y del que dice que arrastra una serie de presupuestos enteramente falsos– que conduce, por sí mismo, a confusiones y enredos evitables desde fuera del mismo, y añade que tanto aquéllos como éstos se equivocan al suponer que sus posturas enfrentadas agotan el abanico de posibilidades (Searle, 1992: 2-3 del original, 17 de la traducción). Tanto unos como otros, tanto dualistas como materialistas, hacen uso de un vocabulario que, según Searle, lleva a unos y otros a idénticos callejones sin salida.

El origen de la concepción tradicional de la contraposición físico/mental lo encontramos en la obra de René Descartes. Sus *Meditationes de prima philosophia* (1641, VI) pueden ser contempladas hoy como el texto en que aparece por vez primera la formulación moderna del problema mente-cuerpo. El texto no deja lugar a la duda: el mundo se halla escindido de tal modo que dos sustancias completamente heterogéneas lo integran y, así, nos integran a nosotros. *Res cogitans* y *res extensa*, dos sustancias con propiedades diversas, excluyentes. La interpretación que Searle (2004a: 14-15) hace del plan-

[27] Más recientemente, Searle ha hablado en términos de *estados* en lugar de *rasgos* del cerebro: "'Consciousness' does not name a distinct, separate phenomenon, something over and above its neurobiological base, rather it names a state that the neurobiological system can be in" (Searle, 2002b: 60).

[28] El autor del artículo citado vuelve a subrayar en su reseña de *The Mystery of Consciousness* (García Valero, 2006: 122) el modo en que Searle entiende que el vocabulario y las categorías tradicionales conducen a los principales callejones sin salida del estudio contemporáneo de la conciencia.

teamiento cartesiano atiende a factores sociológicos: según Searle, Descartes logró temperar los ánimos de la influyente élite teológica de su época, recelosa ante el avance de la nueva ciencia –que adquiría desde determinados puntos de vista visos de desafío a la religión–, al entregar la *res extensa* a los físicos y reservar la inmaterial e inmortal *res cogitans* para la esfera filosófico-teológica. En cualquier caso, y con independencia de las posibles motivaciones que puedan hallarse a la base de los orígenes del dualismo cartesiano, Searle entiende que se trata de una concepción desorientadora, totalmente errónea, de hecho. No obstante, y a pesar de apuntar con sus dardos a las categorías tradicionales, Searle considera tan erróneas como las tesis dualistas de las que surgiera el aparato conceptual tradicional las de los materialistas contemporáneos que hacen uso del mismo, y dedica a la crítica de estas últimas más espacio que a la de las tradicionales doctrinas dualistas.[29] Sin embargo, Searle se ha pronunciado explícitamente no sólo respecto del dualismo cartesiano, sino también respecto del dualismo de propiedades (vid. Searle, 2002b), subrayando que no debe interpretarse su propuesta en semejantes términos. El motivo por el cual pretende definir su postura marcando las distancias con el dualismo de propiedades es el siguiente: mientras muchos filósofos de la mente contemporáneos encuentran pocas diferencias entre el planteamiento de Searle y el propio del dualismo de propiedades (Stich, 1987; Nagel, 1993a; Kim, 1995; Hierro-Pescador, 2006: 79), Searle considera que los dualistas de propiedades han rechazado el dualismo sustancialista sólo como un gesto, porque en su marco teórico la conciencia vuelve a aparecer fuera o por encima del resto del mundo físico, mientras, por su parte, él pretende presentárnosla como un rasgo del cerebro (*a feature of the brain*), un rasgo de *nivel superior* e *irreductible* –ontológicamente–, pero no por ello una *propiedad* independiente o separable –concepción que Searle atribuye a filósofos contemporáneos como Jaegwon Kim–. En resumidas cuentas, Searle (2002b: 58-59) describe el dualismo de propiedades como una doctrina integrada por las tres tesis siguientes: 1) existen dos categorías metafísicas que se excluyen mutuamente y que constituyen toda la realidad empírica: lo físico y lo mental; 2) como los estados mentales no son reductibles a los estados neurobiológicos, son entonces algo *distinto de* y *por encima y además de* los estados neurobiológicos; y 3) los fenómenos mentales no constituyen objetos o sustan-

[29] No comentaremos las críticas de Searle al dualismo. No obstante, es interesante señalar que Searle (2004a: 42-43; 2007b: 175 y ss.) ha considerado con cierto detalle los problemas a los que conducen tanto el dualismo como postura general como el dualismo sustancialista –problemas como el epifenomenalismo o la sobredeterminación causal, relacionados con la causación mente-cuerpo y con la forma de concebir la interacción entre dos esferas ontológicas totalmente heterogéneas–, y que ha tratado de distanciarse particularmente del dualismo de propiedades (vid. Searle, 2002b).

cias separadas, sino propiedades de la entidad completa. Este marco teórico, según Searle, conduce a las paradojas causales propias de enfoques epifenomenalistas (en esencia, la de que la conciencia carezca de poderes causales en un mundo físico causalmente cerrado), al tiempo que al reverso de las mismas: el problema de la sobredeterminación causal, es decir, el de contar con varias causas distintas y suficientes (físicas y mentales) para un único efecto (Ibíd.: 59). Estas paradojas y problemas, entiende Searle, desaparecen tan pronto como abandonamos las habituales distinciones entre categorías metafísicas discretas y dejamos de concebir la conciencia como un *extra* de los procesos neuronales, es decir, como algo separado o además los mismos. Como telón de fondo de la crítica de Searle al dualismo de propiedades encontramos su redefinición de lo físico y lo mental como nociones no opuestas: Searle hablará de niveles de descripción y no de categorías ontológicas, con lo cual prepara el terreno para la comprensión de la conciencia como una de las caras de la moneda (única) de la actividad nerviosa.

No obstante su crítica del dualismo de propiedades y del dualismo cartesiano, Searle hace gala de un realismo mentalista de corte cartesiano al afirmar (Searle, 1992: 13 del original, 28 de la traducción) que cabe aceptar los hechos obvios acerca de nuestras vidas mentales que acogía el cartesianismo sin necesidad de aceptar por ello el aparato conceptual y teórico cartesiano que tradicionalmente ha acompañado al reconocimiento de esos hechos. A pesar, pues, de su intención de sortear la Escila del materialismo y la Caribdis del dualismo (Searle, 2006: 107), Searle considera oportuno el intento de rescatar lo que de rescatable haya en cada extremo.[30] Así, insiste (Searle, 1992: 28 del original, 42 de la traducción), pueden aceptarse los *hechos obvios* de la física sin negar por ello los *hechos obvios* de nuestra experiencia consciente (apuntemos de pasada que en este segundo grupo de *hechos obvios* incluye Searle el de que los estados conscientes tienen propiedades fenomenológicas completamente *irreductibles*), y añade que considerar contradictorios ambos actos de aceptación es un error, un error que deriva de la asunción de una serie de presuposiciones subyacentes al vocabulario tradicionalmente empleado a la hora de abordar el problema de la conciencia.

> "[Searle] believes that it is perfectly consistent with naturalism to hold that the world is entirely composed of physical particles obeying the laws of physics while still maintaining that there are irreducible features of the mind that fit perfectly well into a naturalistic physical ontology" (Sorem, 2010: 32).

[30] "Both dualism and materialism are false, but both are trying to say something true and we need to rescue the true part from the false part" (Searle, 2007b: 170).

Aludíamos en el párrafo anterior a la intención de Searle de sortear el dualismo y refutar el fisicalismo, lo cual habría de sonar extraño siempre que, con Searle, consideráramos la propia distinción entre materialismo y dualismo como formando de algún modo parte de una manera obsoleta de trazar demarcaciones. Sea como fuere, Searle sostiene que el dualismo es difícil de refutar pero inconsistente con nuestro cuerpo actual de conocimientos acerca del funcionamiento del universo –llegando a hablar del dualismo en términos de creencia irracional (Searle, 2004a: 132)–, mientras, por otra parte, dirá del materialismo –materialismo… en términos searleanos– que es fácilmente confutable por cuanto la conciencia existe y todos lo sabemos. Uno de los puntos en los que Searle insiste respecto de este extremo –la distinción materialismo/dualismo– es que es importante tener presente que el dualismo y el materialismo no agotan las posibilidades de cara a enfrentarse al problema de la conciencia.

Apuntábamos que las críticas de Searle a lo que él denomina materialismo en filosofía de la mente son más extensas que las que dirige a las tesis dualistas. Veámoslas con algún detalle. El primer punto a aclarar es que la noción de materialismo en Searle tiene un carácter propio, pues alude con la misma a planteamientos teóricos que reconocen sólo la existencia de lo físico, en tanto ontológica y epistémicamente objetivo, y dejan de lado la subjetividad y su ontología de primera persona (Searle, 1992: 27 del original, 41 de la traducción). En este sentido, podemos entender a Searle como un filósofo materialista sólo en tanto suscribiría la tesis de que el universo está enteramente compuesto de materia –másica y no-másica– y en él todos los procesos no son sino procesos físicos. Sin embargo, el sentido reduccionista al que apuntábamos, el sentido eliminativista del que Searle dota al término –pues, según Searle (2006: 105), el reduccionismo es una forma de eliminativismo–, es el principal objetivo de sus dardos. Por materialismo en filosofía de la mente entiende Searle, pues, la negación de la existencia de una ontología de primera persona, la negación de la existencia de la conciencia como un fenómeno –tal y como Searle defiende– irreductible.

El recorrido de su crítica del materialismo parte de consideraciones históricas. Searle se pregunta cuáles hemos de entender que han sido las derrotas transitadas en la moderna historia intelectual occidental hasta desembocar en el punto actual, es decir, se pregunta cómo o de dónde ha surgido un clima intelectual en el que las tesis materialistas –reduccionistas, eliminativistas– que él considera obviamente falsas suenan plausibles a oídos de tantos. Su respuesta consiste en a) llamar "tradición" al conjunto de planteamientos cuyo desarrollo ha provocado el señalado clima, b) enumerar los rasgos

cruciales de esa tradición, y c) apuntar a cuatro causas históricas conjuntas del *cambio climático* en cuestión.

Será mediante la referida enumeración que Searle (1992: 10-11 del original, 24-26 de la traducción) perfile los contornos de eso que llama en este punto tradición. De este modo, esa tradición consiste, de acuerdo con Searle, en un conglomerado –de carácter a menudo implícito– de tesis, opiniones y suposiciones metodológicas integrado por al menos seis elementos diferenciables. 1) Por lo que al estudio científico de la mente se refiere, la conciencia es de, digamos, exigua importancia: es dable y deseable explicar *la cognición* sin prestar atención a la subjetividad. 2) La ciencia es objetiva, no sólo en el sentido de que trata de obtener resultados, conclusiones, leyes o teorías independientes de prejuicios o sesgos personales, sino en el sentido de que trata con una realidad objetiva: la realidad misma es objetiva. 3) Dado que la realidad es objetiva, el estudio científico de la mente habrá de realizarse desde el punto de vista objetivo o de tercera persona. La objetividad de la ciencia requiere la objetividad de los fenómenos estudiados, lo cual, en el caso de la ciencia cognitiva, significa que el estudio científico de la mente debe atender a hechos objetivamente observables. 4) Desde el punto de vista objetivo de tercera persona la única solución al así llamado problema de las otras mentes pasa por la observación de la conducta. 5) La conducta y las relaciones causales con la conducta son la esencia de lo mental. 6) En última instancia, sólo existen cosas físicas, en el sentido tradicional de «físico», esto es, en un sentido contrapuesto a “mental”.

Searle se propone mostrar que todos y cada uno de los elementos de este conglomerado tradicional son falsos, que ofrecen, en conjunto, una idea incoherente de la realidad. Principalmente, le interesa subrayar que dicho conglomerado no deja espacio para la aceptación de aquello que le parece más obvio: la existencia de la conciencia como fenómeno ontológicamente subjetivo, como fenómeno caracterizado por su ser-para-un-sujeto, esto es, por su existencia para la primera persona. Así, 1) al primer elemento del conglomerado tradicional opone una intuición suya según la cual no hay forma de estudiar los estados y procesos mentales sin atender, explícita o implícitamente, a la conciencia, dado que, en su opinión, no tenemos una noción de lo mental independiente de la noción de conciencia; 2) defiende que no toda la realidad es objetiva, dado que la subjetividad (ontológica) forma parte de la realidad, la cual cabe estudiar tratando de poner límites a la subjetividad epistémica, mas no a la ontológica; 3) sugiere que el estudio científico de la mente no debe perder de vista la perspectiva de primera persona, dado que la ontología de lo mental es la de la primera persona –añade en este punto que

ha de considerarse con cautela la pretendidamente omniabarcante contraposición conductismo/introspección–; 4) plantea que la solución del problema de las otras mentes no depende exclusivamente de la conducta, sino de *una cierta concepción causal del funcionamiento del mundo* en virtud de la cual sabemos que un organismo es consciente porque sabemos, en cierto sentido, cómo funciona; 5) supone que la conducta o las relaciones causales con la conducta no son esenciales para la existencia de fenómenos mentales y que la relación de los estados mentales con la conducta es *netamente contingente* –"purely contingent" (Ibíd.: 23 del original, 37 de la traducción)–: unos y otra, según Searle, pueden existir con independencia –destaquemos en este punto que Searle utiliza aquí la noción de conducta en un sentido laxo, estrecho o macro, dado que la concibe como algo que puede suprimirse con la sola desconexión de eferencias o la mera destrucción del sistema nervioso motor–; y 6) señala que la concepción cartesiana de lo físico es inadecuada, dado que se articula en torno a distinciones conceptuales caprichosas, como la que supone una oposición, que Searle califica de falsa, entre lo físico y lo mental, dado que la física actual descalifica la concepción de realidad física interpretada como *res extensa*, y dado que la afirmación de que la realidad es enteramente física no implica la inexistencia de cosas tales como los subsidios por desempleo –es decir, hechos institucionales, que Searle define como relativos al observador y se muestra dubitativo a la hora de determinar hasta qué punto cabe conceptualizarlos en términos descriptivos o valorativos (vid. Searle 1964; 1969; 1995a)– o la subjetividad (ontológica) de la conciencia. Como apuntábamos, Searle pretende mostrar lo erróneo de estos supuestos, pero, "ante todo y principalmente, no desea admitir el último supuesto (…) tradicional, según el cual la existencia de lo físico genera oposiciones como dualismo-monismo o mentalismo-materialismo. Él acepta que todo, en última instancia, es físico, pero [entiende] que esta admisión no debe generar oposiciones sino inclusiones: la conciencia es sobre todo un fenómeno biológico y lo biológico se incluye en lo físico" (Daros, 2005: 131).

En conclusión, Searle opina que el materialismo contemporáneo en filosofía de la mente es, en último término, tan dualista como el cartesianismo, pues ha adoptado su vocabulario y es presa por ello de los enredos a los que el mismo conduce: los materialistas contemporáneos en filosofía de la mente, dirá, utilizan las categorías del dualismo tradicional en su conato de cortar la tarta de tal modo que sólo una de las mitades (la "física", en detrimento, pues, de la "mental") pueda ser entendida como verdaderamente real. De este modo, el materialismo contemporáneo en filosofía de la mente supone im-

plícitamente un dualismo conceptual al aceptar las nociones de lo físico y lo mental como opuestas y mutuamente excluyentes. Es aquí donde entran en escena las cuatro referidas causas históricas confluentes, pues, según Searle, era de esperar que, una vez aceptado el aparato conceptual cartesiano y transitado el curso de la moderna historia intelectual de occidente, acabáramos por desembocar en el materialismo (que, insistamos, en términos searleanos viene a significar eliminativismo) en la filosofía y las ciencias de lo mental, en una tradición que "nos ciega para los hechos obvios de nuestras experiencias" (Lukomski, 2007: 65) induciéndonos a considerar falsos enunciados evidentemente verdaderos acerca de la mente y la conciencia y haciendo que hipótesis obviamente falsas resulten plausibles (Searle, 1992: 12 del original, 26-27 de la traducción). Aproximémonos brevemente a esas cuatro causas históricas que, en la interpretación de Searle, han venido a conformar esta tradición materialista.

1) En primer lugar, llama Searle la atención sobre el miedo al dualismo cartesiano, una actitud cuyo frecuente contrapunto consiste en la negación de hechos obvios acerca de nuestras vidas mentales, hechos de sentido común cuyo único pecado consiste en sonar cartesianos. En este miedo estribaría, según Searle, el rechazo explícito o implícito de las tesis mentalistas por parte del grueso de la comunidad filosófica y científica, que, siempre según Searle, temerosa de la metafísica cartesiana, se muestra incapaz de disociarla de los referidos hechos obvios, pues persiste la creencia de que al aceptar éstos recaeríamos en aquélla (Ibíd.: 13 del original, 27 de la traducción). Sin embargo, y frente a esta tendencia, reconocer la existencia de la conciencia, propone Searle, no implica ninguna clase de compromiso con el cartesianismo ni supone un riesgo para el teórico interesado en evitar el dualismo.

2) Relacionado con el primer punto se encuentra la ya referida herencia de un vocabulario cartesiano, al cual se hallan asociadas una serie de categorías (las más relevantes de la cuales se ubicarían a ambos lados de la oposición mente/materia) que condicionan nuestro abordaje del problema de la conciencia. La herencia de semejante conjunto de categorías hace además caer, según Searle, en la ilusión de que el espectro tradicional de posturas que encontramos a ambos lados de las oposiciones fisicalismo/mentalismo y monismo/dualismo comprende de hecho el espectro completo de posturas posibles.

3) La tercera pata del banco de la contemporánea tradición materialista es, en la reconstrucción de Searle, una persistente tendencia a la objetivación ínsita en el orbe intelectual occidental desde el siglo XVII y consistente en la equiparación de las nocio-

nes de lo objetivo y lo científico no sólo en el plano epistémico, sino también en el ontológico. Sus repercusiones se hacen notar, según Searle, en el desoír de la subjetividad (ontológica) por parte de la filosofía y las ciencias de lo mental, y en el consiguiente desplazamiento de la atención desde aquélla hacia la objetividad de la conducta.

4) Para finalizar, Searle encuentra el cuarto resorte de la conformación de la tradición materialista contemporánea en *la embriaguez de las grandes profundidades* (Austin, 1956: 179), el fenómeno de hacer ascos a tesis humildes y triviales –precisamente por humildes y triviales– y favorecer propuestas esotéricas, sorprendentes y contraintuitivas.

En resumidas cuentas, Searle encuentra que el vocabulario tradicional arrastra consigo una serie de errores y enredos conceptuales que hacen difícil incluso la interpretación de sus propias tesis, de forma que entiende (vid. Searle, 1999a) que la única salida consiste en abandonar ese vocabulario tradicional.

> "Los materialistas dicen, «la conciencia es sólo un proceso cerebral». Yo digo, «la conciencia es sólo un proceso cerebral». Pero los materialistas quieren decir: la conciencia como un fenómeno irreductible, cualitativo, subjetivo, de primera persona, etéreo, íntimo, no existe. Sólo existen fenómenos objetivos de tercera persona. Pero lo que yo quiero decir es que la conciencia, precisamente como un fenómeno irreductible, cualitativo, subjetivo, de primera persona, etéreo, íntimo, es un proceso que tiene lugar en el cerebro. Los dualistas dicen, «la conciencia es irreductible a los procesos neurobiológicos de tercera persona». Yo digo, «la conciencia es irreductible a los procesos neurobiológicos de tercera persona». Pero los dualistas consideran que esto implica que la conciencia no es parte del mundo físico ordinario sino algo por encima de él. Lo que yo quiero decir es que la conciencia es causalmente reductible pero no ontológicamente reductible. Es parte del mundo físico ordinario y no es algo por encima de él" (Searle, 2004a: 126-127τ).

Searle se propone superar estos enredos conceptuales y estas tradiciones dualistas y materialistas mediante su naturalismo biológico, el quid del cual consiste en –un intento de– tratar a la conciencia como un fenómeno biológico más, como parte del orden biológico natural, definiéndola como un rasgo biológico –y por tanto físico (Searle, 1992; 1995a; 1997a; 1998; 2002b; 2004a)– del cerebro (*a feature of the brain*). La etiqueta "naturalismo biológico" pretende responder a tres cuestiones: la necesidad de contar con una denominación que diferencie su perspectiva de la materialistas y dualistas, la de subrayar que el biológico es el nivel adecuado para la explicación científica de la conciencia, y la de enfatizar que la conciencia forma parte del mundo natural. Searle

(2004a: 113-114; 2007a: 327-328; 2007b: 170-172; 2009: 107-109) presenta el núcleo de su naturalismo biológico mediante las siguientes cuatro tesis.

1) *La realidad y la irreductibilidad de la conciencia*. Los estados conscientes, con su ontología subjetiva, de primera persona, son fenómenos que no cabe reducir eliminativamente mostrando que se trata de una mera ilusión o de *nada más* que actividad nerviosa, porque una reducción de este tipo, una reducción de la subjetividad a hechos de tercera persona, excluiría el aspecto decisivo de la conciencia: su ontología de primera persona.

2) *La causación neurobiológica de la conciencia*. Los estados conscientes son completamente causados por procesos neurobiológicos de nivel inferior en el cerebro y, por tanto, podemos entender la conciencia como *causalmente reductible* a procesos neurobiológicos. La conciencia no es un extra de la neurobiología, no se halla por encima de ella ni es en absoluto independiente de la misma.

3) *La realización de la conciencia en el cerebro*. Los estados conscientes son realizados en el cerebro, son rasgos suyos, propiedades biológicas –y por tanto físicas– del cerebro, aunque propiedades de nivel superior (higer-level features).

4) *El funcionamiento causal de la conciencia*. Los estados conscientes, como rasgos del cerebro y por tanto parte del mundo físico, funcionan causalmente.

Podemos resumir estas tesis señalando que la conciencia es presentada por Searle como un rasgo del cerebro ontológicamente subjetivo, ontológicamente irreductible, causalmente reductible y causalmente eficaz. "El 'naturalismo biológico' propugnado por Searle vendría a reivindicar, a la vez, la irreductible existencia ontológica de los fenómenos mentales y su consideración como productos del funcionamiento neurológico del cerebro" (Hermoso & Chacón, 2000: 172).

La cabal comprensión de estas tesis depende de una atenta consideración de los siguientes extremos: el referido cambio conceptual, la noción de causación y la de reducción. Veamos el tratamiento que Searle ofrece de cada uno de los mismos.

a) Por una parte, las tesis centrales del naturalismo biológico requieren del apuntado cambio conceptual: la incomprensión de estas tesis resulta inevitable si concebimos "físico" y "mental", en línea con la tradición, como términos mutuamente excluyentes. Este viraje conceptual, consistente en concebir "mental" y "físico" como términos no opuestos entre sí, resulta crucial dentro del marco del naturalismo biológico searleano,

dado que en él descansa la posibilidad de comprender la conciencia como un proceso biológico entre otros. Searle (2004a:115) sugiere que la oposición de términos (realidad física frente a realidad mental) que se propone soslayar no proviene sólo de una herencia conceptual, sino que se trata además de una ingenua distinción de sentido común, pero, contra la habitual salvaguarda que Searle ofrece al sentido común, en este caso se trata de una intuición de sentido común errónea, y el error consiste en interpretar esa ingenua intuición de sentido común (la diferencia entre estados mentales y estados físicos) como expresión de una distinción metafísica profunda. Con la denuncia de este error Searle trata de presentar a la conciencia como un fenómeno no superpuesto a la realidad física, sino in-corporado, trabado con nuestra realidad corpórea. Pretende así sortear los tentáculos del dualismo y ofrecer un marco para la comprensión de la conciencia que salve la brecha entre lo físico y lo mental articulándolos sin recluirlos en esferas heterogéneas. En su propuesta, en definitiva, "físico" no significa no-mental y "mental" no significa no-físico (Searle, 1992: 14-15 del original, 29 de la traducción),[31] sino que, al contrario, lo mental en tanto mental es físico en tanto físico (Searle, 2004a: 118) y, así, la ontología de primera persona, la intencionalidad y todo lo que denominamos mental está localizado espacial y temporalmente en el cerebro y es causalmente explicable al tiempo que capaz de actuar como causa (Ibíd.: 117). Searle hablará de una dependencia causal entre cerebro y conciencia al tiempo que propondrá que cerebro y conciencia no son exactamente lo mismo, pues existe (Ibíd.: 130) una palmaria diferencia entre aquellas características irreductibles del mundo que poseen una ontología de primera persona y aquellas que no, lo cual no implica que ambos grupos de características pertenezcan a esferas independientes e incomunicadas de la realidad o que los fenómenos subjetivos sean algo por encima o más allá de los sistemas en los que son realizados: "si bien los fenómenos mentales son causados (…) por la acción cerebral, no se trata empero de una entidad diferente de ést[a]. Así, Searle afirma que los estados y procesos mentales son causados por las operaciones del cerebro y asimismo realizados en la estructura del cerebro y del resto del sistema nervioso central" (Guarino, 2010: 35).

b) Ya un superficial vistazo de las cuatro tesis enunciadas revela que su sentido depende, en gran medida (particularmente por lo que a la segunda y la cuarta se refiere), de la concepción de la causación que tengamos en mente a lo hora de darles contenido.

[31] Searle, cosa no habitual en su forma de discurso, llega a expresarse en términos ciertamente crípticos a la hora de presentar su planteamiento en este punto: "Consciousness *qua* consciousness, *qua* mental, *qua* subjective, *qua* qualitative is physical, and physical because mental" (Searle, 1992: 15 del original, 29 de la traducción).

Sin embargo, Searle, ciertamente, no ha elaborado un marco teórico exhaustivo del tipo de los habituales en filosofía de la ciencia para la presentación de su noción de causación. De hecho, sus tentativas de aproximación a una noción general de causación llegan a sonar triviales y hasta tautológicas, como cuando propone que la noción general de causación es la noción de algo haciendo que otra cosa ocurra (Searle, 2001: 40), o cuando propone que la causación es una relación real entre objetos y eventos en el mundo mediante la cual un fenómeno causa otro (Searle, 1997b: 10). A pesar de ello, y tal y como apuntábamos al hablar del modo en que Searle incluye la teoría atómica de la materia entre los elementos fundamentales de nuestra contemporánea visión científica del mundo, el de Denver esboza un modelo causal para su concepción de la conciencia, un *modelo* que considera de una importancia decisiva de cara a comprender su propuesta.

Atenderemos al señalado modelo tras hacer mención brevemente de la concisa crítica de Searle de lo que entiende como concepción habitual de la causación. Searle considera que la concepción consuetudinaria de causación dificulta la aceptación de sus tesis dado que en la misma la causación queda definida como una relación entre eventos discretos y consecutivos en la que la causa antecede necesariamente al efecto, un prejuicio, dice, que impide contemplar cómo, en efecto, son muchos los casos en los que causa y efecto son fenómenos simultáneos, como en los habituales ejemplos de Searle en los que la conducta de elementos del micronivel es la causa de propiedades en el macronivel. Esto nos da pie para entrar ya brevemente en el señalado modelo, pues la idea de una causación simultánea de micronivel a macronivel es el núcleo del mismo, y, asimismo, uno de los elementos fundamentales del naturalismo biológico. Searle entiende pues como causal la propia constitución de los objetos, y habla de dicha causalidad como *causalidad abajo-arriba*: los elementos del micronivel causan efectos simultáneos en el macronivel, y por esto precisamente aquellos microfenómenos sirven para explicar causalmente (Searle, 2004a: 124) características macro de los sistemas.

> Es esencial al aparato explicativo de la teoría atómica no sólo la idea de que los sistemas grandes están hechos de sistemas pequeños, sino [asimismo la de] que muchos rasgos de los grandes pueden ser *causalmente explicados* por la conducta de los pequeños (Searle, 1992: 87 del original, 99 de la traducción).[32]

La clave de su propuesta está así en una distinción, trazada con unos contornos gruesos y tentativos (es decir, «grande» y «pequeño» no son habitualmente considera-

[32] Cursivas en el original.

dos paradigmas de precisión descriptiva), entre niveles macro y micro de descripción y explicación.[33] Por extraño que resulte, esto es todo lo que resulta de interés del señalado "modelo" por lo que al problema de la conciencia respecta, y no sólo esto, sino que, además de la escasa amplitud del tratamiento de Searle de la noción de causa, la solidez de dicho "modelo" depende enteramente de nuestra comprensión de los ejemplos que Searle utiliza para ilustrar su noción de causación. No es de extrañar, pues, que se hayan alzado voces denunciando la laxitud del tratamiento que Searle hace de esta noción, central en su planteamiento y ampliamente discutida a lo largo de la historia de la filosofía moderna, particularmente, dentro del ámbito de la contemporánea filosofía de la ciencia.[34] Algunas de estas voces (vid., v. g., Pérez Chico, 1999) han señalado explícitamente que Searle no ha aclarado en absoluto qué entiende por causación. Tratemos, no obstante, de analizar algunas de las implicaciones de su "modelo".

Hemos indicado que en la medula del mismo se halla la distinción entre niveles micro y macro de descripción y explicación, y en la comprensión del sentido en que elementos del micronivel causan efectos simultáneos *abajo-arriba* en el macronivel: esta distinción es la que lleva a Searle (1992: 111 del original, 122 de la traducción; 1998: 385) a hablar de la conciencia como una propiedad causalmente emergente. Esta aparentemente inocua distinción, por otra parte, arrastra implícitamente una serie de supuestos que conciernen a nociones problemáticas, como las de emergencia[35] o reducción. El carácter problemático de dichas nociones en este contexto tiene que ver con la propia semántica de las mismas. Así, la noción de propiedad emergente surge en el contexto de la filosofía británica de la segunda mitad del siglo XIX como un intento de res-

[33] Distinción que constituye el núcleo de un movimiento conceptual mediante el cual Searle pretende evitar la posible réplica dualista: "si la conciencia es causada por procesos cerebrales, entonces es algo distinto de los mismos". Según Searle, su concepción de la causalidad, basada en la distinción de niveles y la constatación de la existencia de causación abajo-arriba, es suficiente para eludir semejante réplica: "The fact that there is a causal relation between brain processes and conscious states does not imply a dualism of brain and consciousness any more than the fact that the causal relation between molecule movements and solidity implies a dualism of molecules and solidity. I believe the correct way to see the problem is to see that consciousness is a higher level feature of the system, the behavior of whose lower level elements cause it to have that feature" (Searle 1998: 385).

[34] Una somera idea de la amplitud de la señalada discusión la ofrece la constante publicación de compilaciones, manuales y libros de texto del tipo de Beebee, Hitchcock & Menzies (2009), o la extensión dedicada, en proporción, a la discusión de la causalidad en antologías de filosofía de la ciencia como Lange (2007).

[35] Searle (1992, 1997a, 1998, 2000a) utiliza la noción de propiedad emergente para referirse a algo que es difícil disociar de su noción de causación abajo-arriba –de hecho sus ejemplos son en ambos casos los mismos: la solidez, la liquidez, la transparencia y la conciencia–. "Con emergencia Searle desea significar que el producto (la conciencia) es *total y causalmente explicado* por los elementos integrados del sistema del cual emerge (del cerebro), aunque no es una propiedad de ninguno de los elementos individuales, ni es un agregado de las propiedades de esos elementos" (Daros, 2005: 133; cursivas en el original).

puesta a la disyunción entre la especificad de ciencias como la química o la biología y su reductibilidad a disciplinas consideradas más básicas (O'Connor & Wong, 2002), un intento de respuesta con el que se trató de abrir un espacio intermedio entre las posturas puramente reductivas del mecanicismo y las antirreductivas del vitalismo, un intento destinado a integrar, en dos palabras, antirreduccionismo y fisicalismo: un intento que, como resulta evidente, cuadra a la perfección con el de Searle. De este modo, la noción de emergencia en Searle, estrechamente relacionada con su noción de causalidad abajo-arriba, contiene una defensa implícita del antirreduccionismo al que su planteamiento del problema de la conciencia está orientado: las propiedades emergentes son, según Searle, aquellas novedosas propiedades que determinados sistemas poseen no en virtud del mero agregado de sus partes, sino de su específica integración y articulación, la cual, –según Searle *causalmente*–[36] da lugar en el nivel macro a propiedades del todo que no pueden hallarse en las partes y que, por tanto, escaparían al minucioso escrutinio de eso que denomina "micronivel". En este sentido, la noción de emergencia tal y como es planteada por Searle trae consigo la idea de que la conciencia es una propiedad física del cerebro humano causada por fenómenos neurobiológicos. No obstante, y a pesar de la insistencia de Searle en que su propuesta logra evitar toda suerte de dualismo, la conciencia nos es presentada desde su noción de emergencia como una novedad respecto de los señalados fenómenos neurobiológicos dotada de un estatuto ontológico irreductible a cualesquiera propiedades de los mismos (vid. Moya, 2009: 388).[37]

c) La elucidación de la noción searleana de causación nos ha conducido a las de emergencia y reducción. Respecto de esta última, nuevamente un superficial vistazo a las cuatro tesis en que Searle resume su naturalismo biológico es más que suficiente para advertir que las mismas dependen de qué entendamos por "reducción". Los últimos apuntes sobre emergencia y causación debieran ponernos ya en la pista del enmarañado parentesco que vincula a estas tres nociones en la filosofía de la mente de Searle. Com-

[36] Hacemos esta acotación por cuanto puede considerarse abierta la opción de concebir las relaciones de emergencia y las de superveniencia como causales o como constitutivas (vid., v. g., Pérez Chico, 1999; Montecucco, 2002). Searle (1992: 111 del original, 121 de la traducción) opta por hablar de "causally emergent system features" y por señalar que sólo la noción de superveniencia causal es relevante por lo que a la discusión del problema mente-cuerpo toca (Ibíd.: 125 del original, 134 de la traducción).

[37] Es interesante no señalar que Searle opina que existen direcciones de causación de lo físico a lo mental y viceversa. Entiende pues que lo mental es causalmente eficaz en dos sentidos: un estado mental puede causar no sólo otros estados mentales, sino también determinadas transformaciones físicas. Searle defiende así la existencia de causación arriba-abajo, ejemplificada en estados mentales que causan eventos neurofisiológicos: "If someone says to me 'secrete acetylcholine at the axon end plates of your motorneurons or I will blow your brains out!' I will swiftly do some downward causation, e.g., by trying to raise my arm, which I know will cause the secretion of the acetylcholine. Here the higher order mental state causes the lower order physiological event" (Searle, 1995b: 219).

probemos si explicitar su noción de reducción atenúa o da pábulo a esta sensación de confusión.

Searle habla de diferentes tipos de reducción, pero conviene subrayar desde el principio la importancia que para sus tesis tiene la distinción que traza entre reducciones causales y ontológicas, pues dirá de la conciencia que es reductible causalmente, pero no ontológicamente (Searle, 2004a: 119), lo que vendría a significar que podemos explicar la conciencia causalmente desde la neurobiología pero no decir de ella que no sea *nada más* que neurobiología. Searle propone que usualmente las reducciones ontológicas se siguen de reducciones causales, pero que eso no es posible en el caso de la conciencia. Por otra parte, Searle habla de reducciones eliminativas y no eliminativas, y dice de las primeras que dependen de la distinción entre realidad y apariencia, y que precisamente por eso no son aplicables a la conciencia, porque en ella la apariencia es la realidad. Habremos de precisar en lo que sigue el sentido de estas ideas de Searle, pero consideramos que el núcleo de sus intuiciones al respecto han sido expresadas con elocuencia en diferentes lugares por el propio Searle, de modo que dejaremos que sea él quien introduzca y epitome el tema.

> El punto central de tener el concepto de conciencia es captar las características subjetivas y de primera persona del fenómeno y este punto se pierde si redefinimos la conciencia en términos objetivos de tercera persona (2004a: 130τ).[38]

Presentar con cierto grado de precisión la noción de reducción de Searle requiere atender a la serie de distinciones a partir de las cuales la elabora. El primer extremo a tener en cuenta en este sentido es la distinción epistemológica que Searle traza entre "objetivo" y "subjetivo" en términos de predicados de juicios (Searle, 1995a: 8): un juicio epistemológicamente objetivo es aquél que se muestra independiente de opiniones, actitudes, gustos o sesgos personales y del cual es dable esperar que pueda ser calificado como verdadero o falso por cualquier observador competente asistido por los métodos y procedimientos pertinentes, mientras un juicio espistemológicamente subjetivo es, sencillamente, todo lo contrario ("Schopenhauer fue un bípedo implume" frente a "Schopenhauer fue un filósofo de segunda fila"). Por su parte, la distinción ontológica entre lo objetivo y lo subjetivo, tal y como Searle la establece, está destinada, como he-

[38] Un pasaje análogo puede leerse en Searle (1998): "the whole point of having the concept of consciousness was to have a word to name those subjective experiences (…) [and] where the phenomenon that we are discussing is the subjective experience itself, you cannot carve off the subjective experience and redefine the notion in terms of its causes, without losing the whole point of having the concept in the first place" (Searle, 1998: 386).

mos apuntado ya, a presentar a la subjetividad como un modo de existencia particular: el de la conciencia, caracterizada así como poseedora de un ontología de primera persona en virtud de la cual determinado tipo de cosas existen sólo para el sujeto, esto es, tienen un modo de existencia subjetivo (vid., v. g., Searle, 1992: 94 del original, 106 de la traducción). Este juego de distinciones permite a Searle definir a la conciencia como un fenómeno ontológicamente subjetivo pero pasible de un tipo objetivo de conocimiento (de hecho, Searle propone que caben todas las posibilidades, es decir, que puede tenerse tanto conocimiento objetivo como subjetivo de ontologías tanto objetivas como subjetivas). Otra distinción que es interesante tener en cuenta al aproximarse a la postura antirreduccionista de Searle y que éste considera una distinción más básica o fundamental que las tradicionales distinciones entre lo mental y lo físico (Searle, 1995a: 9) es la que traza entre objetos dependientes e independientes del observador. Parte del segundo conjunto hacen tanto las partículas físicas como los estados conscientes, pues los objetos independientes del observador pueden dividirse en los que tienen una ontología objetiva y los que tienen una ontología subjetiva –por lo que toca a los objetos dependientes del observador, Searle hablará de ontologías objetivas que, sin embargo, dependen de lo que (Searle, 1990a) denomina intencionalidad colectiva: un billete de lotería, como objeto, puede existir con independencia de cualesquiera estados mentales, pero no como un billete de lotería–. Searle niega una entre las posibilidades combinatorias que ofrecen las categorías que propone: la subjetividad ontológica no puede en ningún caso ser dependiente del observador: su existencia no puede serlo, pues de ese modo entraría a formar parte de lo que existe con la misma carta de ciudadanía que los billetes de lotería, los de dólar o los goles en un partido de fútbol. Desde el punto de vista de Searle, pues, la existencia de la experiencia consciente no depende en absoluto de nuestras actitudes hacia ella o nuestras creencias acerca de la misma.

Con estas distinciones en mente[39] pasemos a considerar los diferentes tipos de reducción a los que Searle alude.

1._ *Reducción ontológica*. Se trata de la clase de reducción más fuerte, el tipo de reducción al que, según Searle, aspiran el resto: mediante ella se muestra que ciertos tipos de objetos no consisten en nada sino en objetos de otros tipos (Searle, 1992: 113 del original, 123 de la traducción).

[39] Distinciones que, sinópticamente, presentan a la conciencia como un fenómeno biológico con una ontología subjetiva de primera persona que puede ser conocida tanto objetiva como subjetivamente y cuya existencia no depende en ningún caso de los estados mentales de ningún observador.

2._ *Reducción ontológica de propiedades*. Se trata, según Searle, de una clase de reducción ontológica, pero al nivel de las propiedades y no de los objetos. Un ejemplo de este tipo de reducción lo ofrecería la comprensión de la temperatura de un determinado volumen de gas como *nada sino*, es decir, como nada más que la energía cinética media de sus moléculas.

3. *Reducción teorética*. Este tipo de reducción consiste en el intento de deducir las leyes de una teoría científica a partir de las leyes de otra.

4._ *Reducción lógica*. Este tipo de reducción es tratado por Searle como una moda obsoleta según la cual cabe traducir sin residuo las palabras y las oraciones referidas a un tipo de entidad a las referidas a otro tipo de entidad; de este modo, las entidades mentadas por aquellas palabras y oraciones se mostrarían ontológicamente reductibles.

5._ *Reducción causal*. Con este último tipo de reducción Searle se refiere a una relación entre dos clases de hechos mediante la cual se muestra que la existencia y los poderes causales de la clase de hechos reducidos son totalmente explicables en términos de los poderes causales de los fenómenos reductores (Ibíd.: 114 del original, 124 de la traducción). Searle usa aquí, como en tantas otras ocasiones, la solidez como ejemplo: la existencia de la solidez y sus poderes causales se pueden explicar causalmente por los de las moléculas dispuestas en estructuras reticulares.

A estos cinco tipos de reducción Searle agrega una distinción aplicable a cualquiera de ellos: la referida distinción entre reducciones eliminativas y reducciones no eliminativas. Las primeras muestran que un fenómeno en realidad no existe sino que se trata de una *mera apariencia*. Por su parte, las reducciones no eliminativas no pretenden mostrar que algo, en último término, no existe. Así, por ejemplo, cuando se reducen rasgos de superficie a su sustrato causal no pretende mostrarse que, en último término, dichos rasgos de superficie no existan.

En la propuesta de Searle, como indicábamos, la conciencia aparece como reductible causalmente a procesos neurobiológicos –dado que, arguye el filósofo de Denver, tanto la propia existencia de la conciencia como sus características y poderes causales se explican por los de determinados procesos neurobiológicos– pero ontológicamente irreductible a los mismos. Veamos cuáles son los motivos de Searle para defender esta afirmación.

Según Searle (Ibíd.: 115 del original, 125 de la traducción; 1998: 385-386), a lo largo de la historia de la ciencia se ha dado una tendencia con arreglo a la cual las reducciones causales exitosas condujeron a reducciones ontológicas, pues en virtud, pre-

cisamente, del éxito de aquéllas se procedía a redefinir los términos relativos a los fenómenos reducidos de forma que éstos pasaran a identificarse con sus causas. Pero, nuevamente según Searle, este patrón no puede replicarse en el caso de la conciencia: ella no puede ser objeto de esta clase de reducción, tan siquiera en el contexto de una neurobiología completa y perfecta (Searle, 1992: 116 del original, 126 de la traducción). ¿Por qué no? Searle sigue en este punto una línea que se remonta a Nagel (1974), Kripke (1980), Jackson (1982) y sus argumentos antifisicalistas. Por lo que a la irreductibilidad de la conciencia toca, el de Denver está de acuerdo con las conclusiones que los argumentos propuestos por los autores referidos pretenden alcanzar, pero propone dar una orientación ontológica a las perspectivas que los mismos ofrecieron desde puntos de vista eminentemente epistemológicos: la irreductibilidad de la conciencia es para Searle una cuestión epistemológica sólo de forma derivada, pues tiene primariamente que ver con cómo son las cosas y con qué rasgos reales existen en el mundo (Searle, 1992: 117 del original, 127 de la traducción), es decir, con el particular modo de ser de la conciencia, con su ontología de primera persona. Searle desarrolla su argumentación haciendo uso de un ejemplo, el del dolor, y preguntándose qué hechos reales corresponden con la proposición verdadera "tengo ahora un dolor". Su respuesta: dos hechos: mi experiencia subjetiva de dolor y los procesos neurofisiológicos que la causan. Pero los dos hechos poseen una ontología diversa y autónoma, pues los rasgos esenciales del dolor desaparecerían si pergeñáramos una reducción ontológica que tratara de mostrarnos al dolor como *nada más* que los procesos neurofisiológicos que lo causan. Es en este sentido que Searle afirma que ninguna descripción de hechos de ontologías objetivas alcanza a expresar el carácter subjetivo de la ontología de primera persona.

Pero, ¿qué diferencia a la conciencia de las propiedades ontológicamente reductibles a sus sustratos causales? Las reducciones ontológicas se siguen de reducciones causales al descubrir que un rasgo de superficie es causado por el comportamiento de los elementos del micronivel. Pero Searle, sorprendentemente –sorprendentemente dado que acaba de decirnos que la irreductibilidad de la conciencia viene, en primera instancia, ontológicamente determinada–, asegura que la posibilidad de realizar reducciones ontológicas de otra clase de propiedades mas no de la conciencia es una consecuencia trivial de la pragmática de nuestras prácticas definitorias que no acarrea implicaciones metafísicas profundas (Ibíd.: 122 del original, 132 de la traducción), una consecuencia trivial fundada en el hecho de que lo que nos interesa del resto de aquellas propiedades es distinto de lo que nos interesa de la conciencia. Por ejemplo, en el caso de la tempe-

ratura, al reducirla ontológicamente al movimiento –rotacional, traslacional, vibracional– de las partículas, nos damos cuenta de que, de cara a la comprensión científica y la manipulación técnica del mundo, nos interesan sus causas subyacentes y no su apariencia subjetiva (Ibíd.: 120 del original, 130 de la traducción). Dado este interés en las causas físicas subyacentes, redefinimos el fenómeno en cuestión en términos de sus causas y, dada esta redefinición, la reducción ontológica del fenómeno se sigue como una consecuencia trivial de nuestras prácticas definitorias. La reducción ontológica se sigue de la redefinición, pero las sensaciones subjetivas de frío y calor no desaparecen milagrosamente tras la redefinición: ahora, aunque las experiencias subjetivas de frío y calor existan como existieron siempre, sencillamente comprendemos que correlacionan con la energía interna de un sistema termodinámico, y –propone Searle– podríamos no llevar a término la redefinición pertinente, pero hacerlo comporta evidentes ventajas científico-técnicas al propiciar una mayor comprensión y capacidad de control de la naturaleza dado el aumento de nuestro conocimiento de las causas en ella operantes. Así las cosas, Searle estaría planteando que llevar a cabo la referida redefinición y, por ende, la apuntada reducción, no es más que una cuestión de voluntad y conveniencia: cabe la posibilidad de no hacer nada parecido, pero es conveniente hacerlo. La noción de interés parece alzarse de este modo como uno de los elementos cruciales de la concepción searleana de la reducción ontológica, pues, como hemos visto, esta clase de reducción depende en su propuesta de los intereses del investigador científico. Con todo, estos intereses –es decir, el hecho de que las reducciones ontológicas se muestren en la argumentación de Searle como relevantes para el avance del conocimiento causal y el control técnico de la naturaleza– nada dicen acerca de la existencia de experiencias subjetivas: Searle afirma (Ibíd.: 121 del original, 131 de la traducción) que una redefinición del tipo de la necesaria para llevar a cabo reducciones ontológicas en otros ámbitos podría hacerse, si insistiéramos, en el caso de la conciencia, pero semejante redefinición dejaría sin reducir la experiencia subjetiva del mismo modo que la redefinición del calor en términos de movimientos de las partículas deja sin reducir, es decir, no elimina, las experiencias subjetivas de calor. Esto se debe a que, tal y como Searle presenta las cosas, el objetivo de las reducciones es el de eliminar la subjetividad epistémica en favor de una acendrada consideración de las causas subyacentes de los fenómenos que sean el caso, pero éste es un movimiento que no puede realizarse cuando lo que interesa es, precisamente, la subjetividad, y en este sentido puede afirmarse que la irreductibilidad de la conciencia tiene en Searle que ver con el modo en que realizamos reducciones y no con la forma de existir

de los estados mentales conscientes: "[according to Searle's account] the irreducibility of consciousness is relative only to our pragmatic notion of reducibility" (Garrett, 1995: 211). Detengámonos un instante en este extremo.

Las reducciones ontológicas parten de redefiniciones basadas en el descubrimiento de causas subyacentes. Las posibilidades de una mejor comprensión científica del mundo y una eficacia mayor en el control técnico de la naturaleza se hallan en este punto vinculadas con el intento de alcanzar la verdadera realidad expurgada de subjetividad. Este intento respondería a la necesidad o el interés de predecir, controlar, manipular y conocer con precisión los mecanismos causales operantes en la naturaleza. La reducción sirve en este intento para eliminar los elementos subjetivos de nuestro acceso epistémico al mundo: para eliminar las apariencias. Pero no podemos eliminar lo subjetivo de lo subjetivo, no podemos sustraer las apariencias a las propias apariencias. La distinción entre realidad y apariencia hace acto de presencia mostrando que en el caso de la experiencia consciente la realidad es la propia apariencia y por tanto no puede trazarse dicha distinción. Éste es el motivo por el que Searle ha afirmado que "el patrón de nuestras reducciones descansa en el rechazo de la base subjetiva epistémica para la presencia de una propiedad como parte del constituyente último de esa propiedad (…). La conciencia es una excepción a este patrón por una razón trivial. La razón (…) es que las reducciones que dejan fuera las bases epistémicas, las apariencias, no pueden funcionar para las bases epistémicas mismas. En tales casos, la apariencia es la realidad" (Searle, 1992: 122 del original, 131-132 de la traducción). La conciencia queda excluida del patrón por el que el resto de las reducciones ontológicas están cortadas no en virtud de alguna característica misteriosa de la propia conciencia, sino en virtud de los patrones estándar de reducción con los que contamos (Ibíd.: 124 del original, 133 de la traducción), de forma que la irreductibilidad de la conciencia, frente a la reductibilidad de otras propiedades o rasgos de superficie, no reflejaría ninguna escisión o brecha en la estructura de la realidad, ninguna distinción metafísica profunda, sino que más bien respondería al carácter de nuestras prácticas definitorias.

La exclusión de la conciencia del patrón estándar de reducción, señalábamos, se debe antes a nuestras prácticas que al modo de existir de los estados conscientes. Searle apunta así que dicha exclusión no es debida a ningún tipo de propiedad misteriosa que la conciencia posea. Ya indicamos al comienzo de este apartado que uno de los principales objetivos del naturalismo biológico es el de eludir el halo de misterio en el que habitualmente se ha presentado envuelta a la conciencia, y este objetivo no resulta para

Searle contradictorio con la exclusión de la conciencia del patrón estándar de reducción, pues la irreductibilidad de la conciencia nos es presentada en su filosofía de la mente como una consecuencia trivial de nuestras prácticas definitorias y no como una consecuencia misteriosa de alguna clase de escisión metafísica profunda.

La distinción entre apariencias y realidad es, como hemos visto, uno de los anclajes de la argumentación acerca de la irreductibilidad de la conciencia desarrollada por Searle. Las reducciones ontológicas tienen a su base el interés de dejar de lado las apariencias subjetivas y tomar en consideración fenómenos reales con independencia del carácter particular de su apariencia en la conciencia del observador. Pero, ¿responden a este criterio todos los ejemplos de reducción ontológica legítima propuestos por Searle? El caso del color parece claro. Si dejamos de lado la apariencia consciente no hay en el mundo nada como lo que denominamos colores: sólo fotones, luminancia, iluminancia, reflectividad, reflectancia…. pero nada de colores, pues ellos no son una propiedad de la luz, sino de las experiencias relacionadas con la percepción de la misma: son *aparentes*. La distinción entre apariencia y realidad puede funcionar en este caso. El caso del calor es análogo: si todo animal consciente desapareciera mañana, nada en el mundo correspondería a eso que denominamos calor o eso que denominamos frío. La distinción vuelve a funcionar. ¿Sucede lo mismo con la solidez, la liquidez o la impenetrabilidad? Parece que no: el agua seguirá siendo líquida y exhibiendo los poderes causales propios de un líquido aunque las experiencias conscientes relacionadas con la percepción de la liquidez dejaran de existir. Surge así la pregunta: ¿es la liquidez *real* y el color meramente *aparente*? Si así fuera, la liquidez, atendiendo a la argumentación de Searle, sería ontológicamente irreductible: no se trataría de una apariencia subjetiva, sino de un hecho real, y al reducirla a sus bases causales no estaríamos constriñendo la apariencia por mor de la realidad y la objetividad. En resumidas cuentas, siguiendo esta lógica, la liquidez sería tan irreductible como la conciencia. Pero puede que el "*parece* que no" que empleábamos poco más arriba se muestre más que justo, porque puede que la liquidez exista como una propiedad aparente a escala, por así decir: si midiéramos unos cuantos nanómetros, la liquidez consistiría para nosotros en el comportamiento de las moléculas de los líquidos, y en este sentido, aunque las propiedades macroscópicas de los líquidos (o los sólidos) no puedan ser consideradas meras apariencias, sus causas subyacentes a nivel micro pueden ser tenidas por la liquidez real (ellas explican acabadamente cuanto cabe explicar) y el comportamiento de los elementos del micronivel puede entenderse

así como la verdadera naturaleza de la liquidez[40] a pesar de que consideremos los rasgos de superficie de los líquidos como algo con una entidad diversa de la de las meras apariencias. Lo mismo no ocurriría con la conciencia –no podemos encontrar la verdadera naturaleza de la conciencia *debajo* de la conciencia aparente–, y aquí estriba la diferencia entre reducir aquel tipo de propiedades y reducir la conciencia a sus causas subyacentes: al descender de la conciencia a sus causas ya no encontramos la propia conciencia, ya no encontramos su verdadera naturaleza, su forma de existir, sino sólo sus causas mientras que la conciencia en tanto tal permanecería sólo como rasgo de superficie, como un rasgo que aparecería en un nivel descriptivo superior del sistema que fuera el caso, un rasgo caracterizable en los términos propios de ese nivel. En el planteamiento de Searle, pues, una reducción de la conciencia a sus causas físicas dejará siempre fuera de la descripción el fenómeno mismo de la experiencia consciente, el cual se mostraría así irreductible a su base física.

Searle (Ibíd.: 112-113 del original, 122 de la traducción) indica que la noción de reducción está íntimamente ligada a la de identidad: el reduccionismo, dice, conduce a una relación de identidad, pues cuando una cosa puede ser reducida a otra no es sino porque no es nada sino esa otra. Y en este sentido, dada la reductibilidad causal de la conciencia, Searle (2004a: 124) concede que la conciencia es idéntica a actividad neurobiológica, que es sólo un proceso cerebral, pero matiza su noción de identidad complementándola con su noción de niveles de descripción. Así, afirma, el mismo evento tendría en un nivel características neurobiológicas y en otro características fenomenológicas, y con esto pretende haber alcanzado una concepción de la identidad según la cual cabe identificar procesos conscientes y procesos neurobiológicos y hacer a la vez todo lo contrario al subrayar que los procesos conscientes, con su ontología de primera persona, no son lo mismo que los neurobiológicos. Podemos entender que, nuevamente, en función de nuestros intereses, optaremos por un nivel de descripción u otro de cara a dar cuenta de una y la misma cosa.

¿Qué tipo de relación se da, según Searle, entre las propiedades que hallamos en los diferentes niveles de descripción de esos sistemas de partículas en campos de fuerzas en los que, desde la interpretación searleana de las implicaciones de nuestra cosmovisión científica actual, todo consiste, incluyendo los organismos conscientes? Searle

[40] Esta idea, por otra parte, parece entrar en contradicción con el tratamiento que Searle ofrece de las propiedades sistémicas y emergentes, dado que presenta propiedades como la liquidez como propiedades sistémicas que hallamos a nivel macro. Así, ¿cómo podríamos defender que la verdadera naturaleza de la liquidez ha de buscarse en el micronivel?

rechaza, como hemos visto, una concepción reduccionista de dicha relación. Las propiedades mentales de nivel superior de tales sistemas, entiende, no son ontológicamente reductibles a propiedades de los niveles inferiores. Definirá su postura al respecto como una forma de emergentismo a la que le cuadraría la etiqueta de superveniencia causal (Searle, 1992: 125 del original, 134 de la traducción), que habría de ilustrar la idea combinada de las nociones de emergencia y superveniencia tal y como son empleadas en la filosofía de la mente contemporánea: existen características de los sistemas en tanto todos que no son, al menos no necesariamente, características de las partes componentes de esos todos. La conciencia nos es así presentada como irreductible ontológicamente y, a la vez, como una característica causalmente emergente de determinado tipo de sistemas. Al igual que sucede con propiedades sistémicas como la liquidez, propone Searle, no podemos figurarnos que un sistema haya de dar origen a propiedades mentales conscientes atendiendo meramente a la disposición o configuración espacial de sus partes componentes (cosa que si podría ocurrir con propiedades emergentes como la forma de un objeto), sino que se hace aquí crucial la atención a las interacciones causales entre las partes componentes del sistema en ese nivel micro que darán origen a la propiedad sistémica causalmente emergente (macro) que los estados mentales conscientes constituyen.

> [Consciousness] is a causally emergent property of systems. It is an emergent feature of certain systems of neurons in the same way that solidity and liquidity are emergent features of systems of molecules. The existence of consciousness can be explained by the causal interactions between elements of the brain at the micro level, but consciousness cannot itself be deduced or calculated from the sheer physical structure of the neurons without some additional account of the causal relations between them. (Searle, 1992: 112 del original, 122 de la traducción).

Las interacciones causales entre elementos en el micronivel neurofisiológico *explican* (ἐπιστήμη) la existencia de la conciencia en el macronivel mental, la cual es así concebida como –*siendo* (ὄντος)– enteramente causada por la conducta de elementos o fenómenos biológicos en el nivel inferior. De este modo, superveniencia viene a significar en Searle que identidad en el micronivel (físico) implica identidad en el macronivel (mental), y que un cambio en éste implica necesariamente un cambio en aquél; la naturaleza del micronivel neurofisiológico determina pues enteramente *su* naturaleza mental. Escribimos «su» tratando de ser fieles al marco teórico de Searle, que no pretende realizar multiplicaciones ontológicas innecesarias mediante la introducción de nuevas entidades mentales al estilo del dualismo substancialista, pero que, con éste, niega en último

término que lo mental y lo físico –o las propiedades micro y macro de una y la misma entidad física, de acuerdo con el conato de viraje conceptual que Searle se propone emprender– puedan identificarse: Searle pretende diferenciar claramente la superveniencia causal que propugna de la identificación o reducción de lo mental (de esas propiedades de nivel superior causalmente supervenientes en determinados sistemas biológicos) a lo físico, una identificación ontológica que no puede realizarse porque la conciencia tiene una forma subjetiva de ser, un modo subjetivo de existencia, una ontología de primera persona enteramente diversa de la ontología de lo meramente objetivo –ontología ésta que caracterizaría enteramente, entre otras muchas cosas, a los procesos neurofisiológicos de nivel inferior a los que Searle dice que la conciencia no puede reducirse sino causalmente, esto es, en cualquier caso, nunca ontológicamente, pues, insiste, la ontología de lo mental es, irreductiblemente, una ontología de primera persona (Ibíd.: 95 del original, 107 de la traducción).

## 4._La relación entre intencionalidad, estados mentales conscientes y estados mentales inconscientes en Searle

> *De la mano de la conciencia viene la intencionalidad*.
> (Searle, 1995a: 6 del original, 26 de la traducción).

Tal vez la distinción que más páginas ha acaparado en la filosofía de la mente contemporánea sea la trazada entre la mente intencional y la mente fenoménica. Se trata, ciertamente, de una distinción crucial, dado que sobre la concepción que manejemos de la relación entre ambas gravitará, en una medida nada desdeñable, nuestra concepción global de la mente. ¿Es la conciencia fenoménica una condición necesaria para el surgimiento y, en general, la existencia de algo que, verosímilmente, quepa llamar intencionalidad? Expusimos en la segunda parte de esta tesis las razones por las que entendemos que "no sé, no contesto" es la forma correcta de responder a esta pregunta. En el presente apartado volveremos, sin embargo, sobre las razones por las cuales Searle se inclina por el sí.

Como el título del apartado sugiere, la relación entre conciencia e intencionalidad se halla en el planteamiento de Searle vinculada a su concepción del inconsciente, abiertamente basada, por su parte, en su concepción de la conciencia: el inconsciente aparece en Searle como ontológicamente dependiente de la conciencia. De hecho, no sólo su concepción del inconsciente se asienta sobre sus ideas acerca de la conciencia, sino que

toda su concepción de la mente lo hace. "Los estados mentales se distinguen de los otros fenómenos físicos en que son, o bien conscientes, o bien potencialmente conscientes. Donde no hay acceso a la conciencia, por lo menos en principio, no hay estados mentales" (Searle, 1995a: 228τ). Lo mental y lo consciente se hallan en la filosofía de la mente de Searle íntima, esencialmente vinculados. Sostiene que la conciencia es la noción mental fundamental, y que "de un modo u otro, todas las demás nociones mentales –como intencionalidad, subjetividad, causalidad mental, inteligencia, etc.– sólo pueden ser entendidas completamente como *mentales* a través de sus relaciones con la conciencia" (Searle, 1992: 84 del original, 96 de la traducción).[41] De aquí que Andrew Bailey haya afirmado que, según Searle, "los fenómenos mentales no pueden ser entendidos sin una adecuada comprensión de la conciencia, y esto a su vez no puede lograrse sin prestar atención a nuestra *experiencia* subjetiva" (Bailey, 2004: 180τ).[42] Veamos, antes de pasar a enunciar las razones por las que Searle entiende que lo mental depende de la conciencia, el esquema clasificatorio en el que el autor integra dichas nociones.

Searle (1992: 161 del original, 169 de la traducción) esboza una clasificación bastante concisa de lo que hallamos dentro de nuestros cráneos: ellos albergan cerebros –estrictamente, encéfalos– que pueden encontrarse en diversos estados, los cuales pueden ser conscientes e inconscientes, mentales y no-mentales. Así las cosas, los estados no-mentales,[43] propone Searle, serán siempre inconscientes, pero la conversa no será cierta en todos los casos, dado que, siguiendo su clasificación, cabe la posibilidad de hallar dentro de nuestros cráneos estados clasificables como mentales que no sean (actualmente) conscientes. Cuando Searle habla pues de estados no-mentales está refiriéndose a estados del cerebro que de ningún modo pueden ser conscientes –Searle (1992: 153 del original, 162 de la traducción) denomina no-conscientes a este tipo de estados–, y cuando habla de estados mentales se refiere a estados del cerebro bien conscientes o bien

[41] El énfasis es de Searle.
[42] Cursivas en el original.
[43] Establecer una frontera estricta y cuidadosamente delimitada entre lo mental y lo no-mental, como hemos sugerido ya, puede resultar una labor desalentadora. Searle no se complica trazando una línea divisoria demasiado precisa: da por consabido qué sea lo mental y utiliza para ilustrar eso que, supuesta y tácitamente –segunda acepción de dicha entrada en la vigésima segunda edición del diccionario de la lengua española de la RAE–, todos sabemos de antemano ejemplos cotidianos y familiares, presentados desde un realismo mentalista ingenuo dentro del marco conceptual del sentido común y la psicología popular. Por su parte, el ejemplo de fenómeno neurofisiológico no-mental usado por Searle es el de la "mielinización de los axones en [el] sistema nervioso central" (Searle, 1992: 154 del original, 162 de la traducción). La insistencia del de Denver en la naturaleza neurofisiológica de lo mental puede apreciarse en fragmentos como el siguiente: "si los eventos exteriores al sistema nervioso central tuvieran lugar, pero nada sucediera en el cerebro, no habría eventos mentales" (Searle, 1984b: 19τ).

capaces de acceder a la conciencia, de modo que deja abierta una privativa puerta a la existencia de estados mentales inconscientes.

¿Estados mentales inconscientes? ¿De qué tipo? El inconsciente searleano está poblado por estados mentales que pueden, *en principio*, llegar a ser conscientes, y la mente searleana –el conjunto de los estados neurofisiológicos a los que está dispuesto a calificar de mentales, aquéllos de los que entiende que puede afirmarse que tienen propiedades mentales– consiste en un subconjunto discreto de estados del cerebro que, o bien son ya –esto es, actualmente– conscientes, o bien pueden llegar a serlo; estos últimos serían en su planteamiento los únicos legítimos peticionarios con opciones de fungir como "estados mentales inconscientes". Al defender esta restringida extensión de «inconsciente» Searle estaría enfrentándose a planteamientos que postulan la existencia de estados mentales *profundamente inconscientes* o *en principio* inaccesibles a la conciencia, es decir, estados mentales de los que el individuo de ningún modo podría llegar a ser consciente, como los postulados en gran cantidad de teorías y modelos en ciencias cognitivas (piénsese, por ejemplo, en la psicolingüística de inspiración chomskiana o en la psicosemántica fodoriana). Este enfrentamiento se halla inserto en un marco de desacuerdo más general entre los planteamientos de Searle y los habituales en ciencias cognitivas, pues la negación de la posibilidad de la existencia de estados mentales *profundamente inconscientes* o *en principio* inaccesibles a la conciencia no es sino un aspecto de la reticencia de Searle a dividir lo mental en fenomenología e intencionalidad como si de esferas en cierto sentido autónomas se tratara: él considera que la conciencia y la intencionalidad están más vinculadas de lo que el grueso de la teoría y la práctica en ciencias cognitivas desearían. De este modo, juzga que una buena teoría de la intencionalidad no podría ser correctamente elaborada sin tomar a la conciencia en consideración, pues a pesar de que entiende que pueden existir estados conscientes no-intencionales, cree que no cabe decir lo mismo al revés, esto es, que no cabe afirmar que existan estados intencionales que no puedan ser conscientes. Este es el motivo por el cual podemos hablar de su concepción del inconsciente como una concepción de superficie o un modelo de mentalidad inconsciente accesible, y contraponerla a una concepción del inconsciente profundo o modelo de mentalidad inconsciente inaccesible para la mente consciente. Así, como apuntábamos, el rechazo searleano del inconsciente profundo puede ser entendido como un aspecto de su concepción de las relaciones entre la mente intencional y la mente consciente o fenoménica. Dicho rechazo no debe ser comprendido, no obstante, como parco en consecuencias, pues de él depende la conexión

que Searle traza entre mente –en general– y conciencia: si las razones que Searle aduce en favor de dicho rechazo se mostraran deleznables, el núcleo de su planteamiento inseparatista respecto de la intencionalidad habría fracasado. El soporte "argumental"[44] de dicho rechazo lo constituye el Principio de Conexión, el cual, como vimos, postula que "la adscripción de un fenómeno intencional inconsciente a un sistema implica que el fenómeno es en principio accesible a la conciencia" (Searle, 1990b: 586τ). Habiéndonos ocupado ya en el capítulo anterior del modo en que Searle presenta y defiende dicho principio, un principio que, entendemos, no constituye sino una homilía –antes que una argumentación– en favor de una concepción de la intencionalidad según la cual ésta no puede darse en ausencia de conciencia fenoménica, pasaremos ahora a exponer las derivas de la noción searleana de Trasfondo.

### 4.1._Derivas del Trasfondo searleano

La noción de Trasfondo en la filosofía de la mente de Searle es ciertamente una noción amplia, lo cual es más que comprensible, pues Searle reserva para la misma tareas nada humildes –como, por ejemplo, dotar de *estructura* a toda nuestra vida consciente (Searle, 1999c: 7), permitir que tenga lugar la interpretación lingüística (Searle, 1995a: 132) o, en general, hacer posible toda representación (Searle, 1983: 143 del original, 152 de la traducción)–. Dedicaremos las siguientes páginas a presentar esta problemática noción tal y como fuera sucesivamente expuesta en diferentes escritos desde la primera sistematización de la misma en *Intentionality. An Essay in the Philosophy of Mind* (Searle, 1983: cap. 5). Atenderemos pues en ellas a la evolución de la concepción searleana del Trasfondo, una evolución marcada por la inclusión en el seno de la filosofía de la mente de Searle del referido principio de conexión, núcleo cardinal de la crítica searleana de la noción de inconsciente propia de la ortodoxia cognitivista, dado que dicho principio "viene a negar uno de los presupuestos básicos de la ciencia cognitiva" (Hermoso & Chacón, 2000: 172): el de la existencia de fenómenos mentales inconscientes e inaccesibles por principio a la conciencia.

Pueden hallarse antecedentes de la noción searleana de Trasfondo en el trabajo de filósofos que influyeron directa o indirectamente en Searle, como Grice (primero profe-

[44] Como veíamos en el capítulo anterior, Searle no tendría reparos a la hora de admitir la pertinencia de este entrecomillado.

sor suyo en Oxford y luego colega en Berkeley) o Strawson padre (del que también fuera alumno en Oxford), pero también en el de Wittgenstein o Ryle. No pretendemos analizar en profundidad los paralelismos que cabría trazar entre las ideas de los señalados autores y la hipótesis searleana del Trasfondo, pero consideramos que resultará interesante hacer sumaria referencia a algunos "aires de familia" antes de ocuparnos explícitamente de *los lugares* que la noción de Trasfondo ha ocupado en la filosofía de la mente de Searle.

El primer antecedente resulta quizá el más conflictivo, pero también el único que Searle cita expresamente (Searle, 1992: 177 del original, 183 de la traducción), aunque de un modo genérico y un tanto elusivo. Se trata de Wittgenstein, pero, como sugeríamos, Searle no indica qué ideas en concreto deberíamos entender que Wittgenstein anticipó en la línea de su noción de Trasfondo, sino que se limita a decir que puede entenderse que gran parte de la labor teórica desarrollada por el último Wittgenstein es en buena medida sobre el Trasfondo (Ibíd.). Auque encontramos que la alusión es en muchos puntos discutible, puede que Searle tuviera en mente ideas del (así llamado) segundo Wittgenstein tales como los famosos "juegos de lenguaje", las "formas de vida" (noción equiparable al *Trasfondo local* de Searle), la "historia natural del ser humano" (equiparable al *Trasfondo profundo* de Searle)[45] o las "proposiciones marco"[46] o bisagra ("hinge propositions") de *On Certainty* (vid., v. g., Wittgenstein, 1969: §35, §279, §334, §337, §340, §341, §343, §421 o §455).

Respecto de la influencia de Ryle, se trata principalmente de una famosa pareja de giros que Searle toma prestada (no cita a Ryle, probablemente por considerarlo un gesto innecesario dada la fama de la señalada pareja) en su primera descripción del Trasfondo: la contraposición ryleana entre saber-qué y saber-cómo. Veamos cómo la usa: "In order that I can have the Intentional states that I do I must have certain kinds of know-how: I must know how things are and I must know how to do things, but the kinds of 'know-

[45] Para cotejar estas dos parejas de nociones compárese Searle (1983: 143-144 del original, 152-153 de la traducción) con Wittgenstein (1953: §25 y §415). La distinción searlena es presentada en los siguientes términos: "A minimal geography of the Background would include at least the following: we need to distinguish what we might call 'deep Background', which would include at least all of those Background capacities that are common to all normal human beings in virtue of their biological makeup –capacities such as walking, eating, grasping, perceiving, recognizing, and the preintentional stance that takes account of the solidity of things, and the independent existence of objects and other people– from what we might call the 'local Background' or 'local cultural practices', which would include such things as opening doors, drinking beer from bottles, and the preintentional stance that we take toward such things as cars, refrigerators, money and cocktail parties" (Searle, 1983: 143-144 del original, 152-153 de la traducción).

[46] Traducimos aquí la expresión "framework propositions", propuesta en un famoso artículo por el conocido experto en Wittgenstein Robert J. Fogelin (Fogelin, 1985: 5).

how' in questions are not, in these cases, forms of 'knowing-that'" (1983: 143 del original, 152 de la traducción). También, por otra parte, la "concepción cognitivista tradicional" (Ibíd.: 150 del original, 159 de la traducción) a la que Searle opone su noción de Trasfondo es en muchos sentidos similar a la "leyenda intelectualista" que Ryle (1949) ataca, precisamente, haciendo uso de su distinción entre 'knowing-how' y 'knowing-that'.

Otro antecedente puede hallarse, como señalábamos, en Grice, que en su artículo "Logic and conversation" (Grice, 1975) introduce la noción de "implicatura conversacional" para hacer referencia a significados que se hallan implícitos en la conversación sin formar parte de los significados convencionales asignados a las proferencias que sean el caso. Searle, por su parte (en un artículo del mismo año titulado "Indirect speech acts"), explorando los límites de la teoría de Grice constata que las condiciones de indeterminación que afectan a las implicaturas conversacionales griceanas requieren de un recurso explicativo que dé cabida a "información de trasfondo mutuamente compartida por el hablante y el oyente" (Searle, 1975: 61τ), una información de trasfondo que, en paralelo con el desarrollo posterior de su noción de Trasfondo, no sólo incluiría contenido lingüístico.

Un último antecedente del trabajo de Searle sobre el Trasfondo puede encontrarse en la noción de presuposición (*presupposition*) desarrollada por Strawson padre (vid. Strawson, 1950; 1952), de la que Searle no hace uso en su trabajo en filosofía de la mente, sino sólo en el que desarrollara previamente en filosofía del lenguaje[47] (vid. Searle, 1969: 126).

Centrándonos ya en el desarrollo de la noción de Trasfondo y, con ella, en el desarrollo de la concepción del inconsciente en la filosofía de la mente de Searle, comenzaremos por el principio, es decir, por la primera formulación sistemática de dicha noción en Searle (1983). Resulta interesante destacar en primer término que Searle usará en este texto las nociones de intencionalidad y representación como sinónimas:[48] que un

[47] Disciplina filosófica que Searle pasaría a comprender como una rama de la filosofía de la mente: "I belive that the philosophy of language is a branch of the philosophy of mind" (Searle, 1981a: 720); "On my view, the philosophy of language is a branch of the philosophy of mind" (Searle, 1992: xi del original, 11 de la traducción). Searle se aviene así a la extendida –y, entendemos, bien fundada– corriente de opinión según la cual, si bien tras el giro lingüístico la filosofía del lenguaje fue el centro de gravedad del quehacer filosófico, la filosofía de la mente habría venido a desbancarla (vid., v. g., Gomila Benejam, 2007: 196).

[48] Dos años antes de que apareciera el libro de Searle al que estamos haciendo referencia, en un artículo en el que ya hace mención de dicho libro (en nota al pie y presentándolo como de próxima aparición),

estado sea intencional implica que el mismo es representacional. La particularidad del Trasfondo en este marco teórico consiste en que el mismo es presentado en esta primera aproximación como una serie de capacidades no representacionales que posibilitan la representacionalidad mental en general: sobre ellas se asienta, según Searle, el carácter representacional de la mente. La importancia de la noción se hace así evidente a la luz de la faena encomendada a ella por Searle. En esta primera formulación de la hipótesis del Trasfondo el mismo nos es presentado, pues, como el lecho preintencional de la intencionalidad: Searle dirá (1983: 143 del original, 152 de la traducción) que los estados intencionales sólo tienen las condiciones de satisfacción que tienen y –por tanto– sólo son los estados intencionales que de hecho son, contra (*against*) un Trasfondo de habilidades que no son ellas mismas estados intencionales.[49] Los estados intencionales no flotan en el vacío, ni son lo que son aisladamente, ni funcionan como funcionan independientemente. La propuesta de Searle (1983) recoge esta idea en dos dimensiones, pues, además de la señalada orientación –por así decir– vertical, según la cual todo estado intencional se alza sobre un lecho no intencional, la individualización de cada estado intencional tiene lugar asimismo en un plano horizontal de conexiones con otros estados intencionales. Esta segunda dimensión horizontal corresponde a lo que Searle denomina "Red de intencionalidad", noción con la que quiere significar que cada estado intencional determina sus condiciones de satisfacción y es el estado intencional que es sólo en relación con gran cantidad de otros estados intencionales (Searle, 1983: 141 del original, 150 de la traducción; 1984b: 68). La Red sustentaría holísticamente el contenido y las condiciones de satisfacción de cualquier estado intencional. Para tener una creencia, por ejemplo, es necesario tener muchas otras creencias: para que mi creencia "Lisboa está en Portugal" sea efectivamente la creencia que es he de tener asimismo muchas otras creencias relacionadas con la geografía, la geopolítica, etc. Pero el mutuo apoyo entre estados intencionales sería insuficiente porque, según Searle, la intencionalidad (de los elementos de la Red) no se autointerpreta ni se autoaplica. Si tratamos de hacer explícitas todas las creencias necesarias para la existencia y el apropiado funcionamiento de mi creencia según la cual Lisboa está en Portugal, acabamos por desembocar en márgenes de lo mental que ya no pueden considerarse intencionales sino más

Searle (1981a: 721) había afirmado que la clave para la comprensión de la intencionalidad es la representación y la clave para la comprensión de la representación son las condiciones de satisfacción (las condiciones de satisfacción de mi creencia según la cual llueve consistirían en que, efectivanmente, lloviera).

[49] La traducción española contiene en este punto una grave errata, pues obvia el "not" en la frase "capacities that are not themselves intencional states" (Searle, 1983: 143 del original, 152 de la traducción).

bien relacionados con el modo en que nos desenvolvemos en el mundo, con un lecho de capacidades, habilidades, prácticas y supuestos preintencionales necesarios para dar cuenta del funcionamiento de toda representación. Estas condiciones preintencionales de toda intencionalidad no son ellas mismas, según Searle, estados intencionales ni parte de las condiciones de satisfacción de los estados intencionales. Sin embargo, sostiene Searle en su primer trazado de los lineamientos para una teorización de la intersección entre la mente consciente y la mente inconsciente, los estados intencionales que conforman la Red, aun cuando formen parte de o determinen las condiciones de satisfacción de estados intencionales actualmente conscientes, pueden perfectamente permanecer ellos mismos inconscientes. Searle tendrá en mente está idea cuando, posteriormente (Searle, 1992), critique su primera aproximación a la teorización de la señalada intersección refiriéndose a ella como a una concepción de la mente como inventario de estados mentales.

Searle (1983; 1984b) plantea pues que siguiendo los hilos de la Red acabamos dando con el Trasfondo, de forma que entiende que determinar el contenido intencional de un estado mental requiere de algo más que contenidos intencionales que remitan unos a otros: en esta primera etapa de la elaboración searleana de la hipótesis del Trasfondo la intencionalidad de lo mental no se sostiene por o se alza sobre sí misma, sino que descansa en último término sobre unos cimientos preintencionales que constituyen la precondición de la intencionalidad y del funcionamiento de los estados intencionales. A esos cimientos es a lo que Searle se refiere cuando usa el vocablo «Trasfondo», y habla del modo en que los integrara en su filosofía de la mente en términos de *descubrimiento* (Searle, 1991b: 290). Según esto, Searle no habría postulado la necesidad de un Trasfondo preintencional de toda intencionalidad: lo habría descubierto.

El Trasfondo, tal y como Searle lo presenta en esta primera aproximación (Searle, 1983), y a diferencia del modo en que aparecerá en Searle (1992), sustenta la Red de intencionalidad y se diferencia claramente de ella. Ésta la hallamos constituida por estados intencionales, aquél por una serie de "elementos"[50] que Searle presenta como pre-

[50] Searle habla de los "elementos" que integran el Trasfondo con mucha amplitud: hablará de *know-how*, de un conjunto de *saber-cómos* (cómo son las cosas, cómo hacerlas), de supuestos o suposiciones –tácitas, virtuales, implícitas, hemos de entender, pues Searle trata de eludir cualquier viso representacional y, según sostiene (Searle, 1983: 157 del original, 166 de la traducción), el propio leguaje, con su carácter intencional, dificulta la tarea de explicitar los elementos no intencionales de lo mental–, de presuposiciones, destrezas, hábitos, posturas, actitudes, habilidades, capacidades, y prácticas –"assumptions", "presuppositions", "skills", "habits", "stances", "attitudes", "abilities", "capacities" y "practices"–. Searle dice que resulta difícil caracterizar los "elementos" del Trasfondo sin hacer uso de nociones intencionales que hagan pensar en dichos "elementos" como si fueran ellos mismos intencionales, que es justamente lo que

condiciones de la intencionalidad: el funcionamiento del Trasfondo no es intencional, pero él es, sin embargo, necesario para el funcionamiento intencional de la mente.

El Trasfondo consistiría pues en una serie de capacidades no intencionales que hacen posible la intencionalidad de la mente: "the Background provides a set of enabling conditions that make it possible for particular forms of Intentionality to function" (Searle, 1983: 157 del original, 166 de la traducción). Pero, ¿cómo proporciona esas condiciones posibilitadoras? Al igual que en escritos posteriores, Searle caracteriza el trabajo realizado por el Trasfondo en términos causales: dice (Ibíd.: 158 del original, 166 de la traducción) que el Trasfondo funciona causalmente, pero que se trata de un tipo no determinante de causación que cabe caracterizar indicando que el Trasfondo proporciona condiciones necesarias pero no suficientes para comprender, pensar, creer, desear, etc., y que en este sentido se trata de posibilitar y no de determinar.[51]

Searle (Ibíd.: 144 del original, 153 de la traducción) admite, a pesar de la confianza en su hipótesis sobre la existencia del Trasfondo –como sustrato preintencional condición de toda intencionalidad– que traslucen las anteriores referencias, que no sabe cómo demostrarla de forma apodíctica y concluyente. En semejantes circunstancias, ante la imposibilidad en que se halla para formular argumentos demostrativos, concede que lo mejor que puede ofrecer al lector son los resultados de una serie de inquisiciones, los cuales, entiende, habrán de bastar para movilizar el asenso. Destacan en la labor de Searle durante esta primera etapa de elaboración de la hipótesis del Trasfondo dos áreas en las que desarrolla las señaladas inquisiciones, dos intuitivos pilares que seguirán sustentando su hipótesis a pesar de los cambios sufridos por ésta entre *Intentionality* (1983) y *The Rediscovery of the Mind* (1992) –obra en la que vuelve sobre el primero de los mismos, del que pasamos a ocuparnos a renglón seguido–: se trata de la comprensión del lenguaje y la adquisición y posesión de habilidades físicas.

Por lo que al primero de los señalados pilares toca, cabe indicar que los rudimentos a partir de los cuales elaborará Searle su hipótesis del Trasfondo se hallaban ya pre-

---

pretende evitar. Es por este motivo que recurre a nociones que él mismo considera oximorónicas ("suposiciones preintencionales"), y es asimismo por este motivo que, de entre todos los términos que utiliza para describir el Trasfondo, los que más satisfactorios le resultan son los de capacidades y prácticas, pues entiende que tienen menos carga representacional que el resto (Ibíd.: 156 del original, 165 de la traducción).

[51] Entendemos que la última frase puede resultar gramaticalmente desconcertante y poco informativa, pero, por una parte, es prácticamente todo lo que Searle nos dice al respecto y, por otra, es prácticamente una cita literal del siguiente pasaje: "The Background functions causally but the causation in question is not determining. In traditional terms, the Background provides necessary but not sufficient conditions for understanding, believing, desiring, intending, etc., and in that sense it is enabling and not determining" (Searle, 1983: 158 del original, 166 de la traducción).

sentes en un artículo que publicara a finales de la década de los setenta, un artículo intitulado "Literal meaning" (Searle, 1978). En él Searle invoca, por así decir, una "proto-hipótesis" del Trasfondo para explicar el significado literal de las oraciones que, en todo caso, parecen carecer del contexto, las aclaraciones, instrucciones o enmiendas pertinentes. Searle argumenta que los nudos significados de cada uno de los términos que conforman una proferencia particular son piezas insuficientes para reconstruir el significado literal (más tarde extenderá esta idea al significado que intenta comunicar el hablante e incluso a formas no lingüísticas de intencionalidad). Ante una laguna como ésta, Searle apela a una serie de supuestos de Trasfondo implícitos –"background assumptions" (Searle, 1979: 117)– que harían posible el significado literal y, asimismo, innecesaria la labor de aquellas aclaraciones, enmiendas e instrucciones contextualizadoras explícitas. La línea trazada en este punto por este artículo pionero no ha sido desplazada significativamente por el posterior desarrollo de la filosofía de la mente de Searle. Tanto en él como en los escritos searleanos acerca del Trasfondo que le sucederán, la argumentación es en esencia –mas con cambios de acento y reenfoques: en aquel primer artículo hablaba Searle solamente del lenguaje; en escritos posteriores pasará a hablar del Trasfondo en relación con la noción más general de representación– la misma y puede resumirse en pocas palabras: supuestos implícitos –y aun difícilmente explicitables de forma acabada– de Trasfondo se hacen necesarios para el funcionamiento de toda forma de intencionalidad, que requiere, de hecho, y tal y como Searle argumentará con posterioridad al artículo citado, de un conjunto de capacidades de Trasfondo que no son en sí mismas intencionales. En otras palabras: el éxito de cualquier forma de representación sólo se hace posible, según Searle, gracias al sustento ofrecido por una serie de capacidades no representacionales a las que denomina conjuntamente Trasfondo. En el caso del argumento a favor de la existencia del Trasfondo basado en la comprensión de oraciones, Searle propone que las capacidades de Trasfondo fijan diferentes interpretaciones para el mismo significado literal. Por ejemplo, entendemos la acción designada con el verbo «construir» de diferente modo en las oraciones «Andrés construye un lego» y «Andrés construye la Línea Maginot» (sin ir más lejos, contamos con que una cosa debe hacerla con sus propias manos y la otra, incluso en el supuesto de que el propio Andrés la construyera él solito, ayudándose de máquinas pesadas). Searle propone que el significado literal no es por sí mismo suficiente para fijar estas interpretaciones y que las cláusulas en términos de otros contenidos intencionales que podríamos sumar en aras de la precisión serían infinitas e insuficientes en cualquier caso para fijar la interpretación

de la oración que de hecho realizamos o acabamos por realizar, porque cada nueva cláusula podría nuevamente ser interpretada de diferentes modos (siempre cabrá interpretar de forma descabellada cada una de esas nuevas cláusulas) y, crucialmente, porque en ningún caso ellas mismas se autointerpretarían. Un Trasfondo no intencional de capacidades, prácticas y habilidades determinadas es lo que Searle entiende que resuelve el problema: en él se detiene la cadena de intencionalidades indeterminadas y sin él, sin ninguna presuposición de trasfondo, es de hecho imposible fijar una interpretación determinada.

El otro foco de evidencia presentado en Searle (1983) atañe, como indicábamos, a la adquisición de destrezas físicas. Searle ilustra este extremo mediante el ejemplo del desplazamiento de lo intencional a lo no-intencional en el tránsito del noviciado al desempeño experto de la habilidad de esquiar. En un primer momento, según Searle, el aprendiz sigue instrucciones explícitas y dotadas de forma representacional, las cuales van abriéndose paso hacia un substrato no representacional y no explícito conforme el esquiador va adquiriendo pericia. Las reglas e instrucciones que en su momento fueron explícitamente seguidas según su forma representacional acaban por resultar superfluas (Searle, 1983: 150 del original, 159 de la traducción): el experto deja de pensar explícitamente en las reglas que podría decirse que sigue en su desempeño. Cuando esas reglas e instrucciones se mueven progresivamente hacia el Trasfondo no avanzan hacia el mismo manteniendo la forma de contenidos mentales inconscientes pero aun así representacionales, sino que, mediante la práctica, se transforman en una serie de disposiciones y capacidades no-representacionales. El esquiador experto no sigue inconscientemente las reglas e instrucciones que en el momento de su aprendizaje siguió conscientemente, sino que, dice Searle, simplemente esquía. Así, cuando esas capacidades entran a formar parte del Trasfondo, ya no tienen la forma de reglas tal y como solemos figurárnoslas (es decir, como proposiciones), sino que se convierten en capacidades, actitudes, posturas o habilidades no-representacionales. Aquellas reglas e instrucciones explícitas se hunden o retiran (*recede*) hacia el Trasfondo. Con esto cree Searle enfrentarse a los hechos de un modo más económico y cabal que el que pergeñaría la que denomina "concepción cognitivista tradicional", según la cual –y tal y como Searle la entiende– habríamos de interpretar que al ejercicio de destrezas físicas como la que Searle toma aquí como referencia subyace la actividad de una gran cantidad de representaciones mentales inconscientes (Ibíd.: 151 del original, 160 de la traducción).

Un aspecto que merece la pena destacar de esta primera formulación de la hipótesis searleana del Trasfondo es la coherencia de la misma con la postura internalista que el de Denver viene defendiendo. Según Searle (Ibíd.: 154 del original, 163 de la traducción) es un hecho empírico que sin nuestra particular constitución biológica y sin el peculiar conjunto de relaciones sociales en el que nos hallamos inmersos no tendríamos el Trasfondo que tenemos. Sin embargo, este hecho empírico, entiende, no debe ensombrecer el hecho lógico según el cual no existe ninguna necesidad (lógica) que haga que esto deba ser así. Es decir, Searle piensa que el hecho de que tengamos un Trasfondo preintencional concreto, y asimismo un conjunto determinado de estados intencionales, no requiere *lógicamente* que nos hallemos bajo las referidas relaciones bio-sociológicas. Tanto el Trasfondo como cualquier estado intencional, propone Searle, dependen empírica pero no *lógicamente* de un determinado conjunto de relaciones con el mundo y una determinada historia biológica. La educción de la empiría a la necesidad lógica, entiende, resulta injustificada en este punto. El planteamiento internalista general que Searle defiende puede resumirse diciendo que no existe una relación lógica entre el contenido mental y el mundo, que pueden interpretarse desde este punto de vista como independientes. Lo que Searle estaría proponiendo es que esto mismo sucede con el Trasfondo.

Veamos, para acabar con nuestro repaso de la primera formulación searleana de la hipótesis del Trasfondo, cómo quedó dicha noción definida en el texto del que hemos venido ocupándonos:

> The Background, therefore, is not a set of things nor a set of mysterious relations between ourselves and things, rather it is simply a set of skills, stances, preintentional assumptions and presuppositions, practices, and habits. And all of these, as far as we know, are realized in human brains and bodies. There is nothing whatever that is "transcendental" or "metaphysical" about the Background, as I am using the term (Ibíd.).

Antes de pasar a ocuparnos del modo en que la noción de Trasfondo da un giro en Searle (1992) con la inclusión del principio de conexión haremos mención de una dificultad con la que, según el propio Searle, topa esta primera formulación de su hipótesis, una dificultad que alguno de los críticos de su primera noción de Trasfondo no han dejado de señalar. Se trata de la dificultad de definir los "elementos" que integran el Trasfondo. La última cita y la penúltima nota al pie recogen los términos que Searle utiliza para describir esos "elementos" y apuntan a los motivos a los que alude al explicar las

dificultades con las que se encuentra a la hora de aclarar qué es exactamente ese Trasfondo preintencional del que habla. Como indicábamos en la señalada nota al pie, Searle (1983: 157 del original, 166 de la traducción) alega que es el propio leguaje el que dificulta la tarea de caracterizar los segmentos no intencionales de nuestra vida mental: al tratar de *representar* lo que el Trasfondo es, el propio carácter representacional del lenguaje, dice, puede hacernos pensar que también el Trasfondo funciona de forma representacional o está integrado por instancias que pueden ser cabalmente entendidas como intencionales o representacionales. Searle achaca pues sus dificultades a la propia naturaleza del lenguaje: éste no sirve para transmitir plenamente el carácter de preintencional de los referidos segmentos de lo mental más que de un modo indirecto que, además, puede hacernos caer en la trampa de pensar que nuestras representaciones –lingüísticas acerca del Trasfondo– son sobre representaciones. Searle recurre a un ejemplo tomado del uso del lenguaje cotidiano para ilustrar las apuntadas dificultades. Nos pide que nos figuremos una situación que ejemplifica la sorpresa que puede experimentarse al levantar un objeto que uno "consideraba" más pesado de lo que resulta ser. Dice que describir la situación en esos términos es ya engañoso, aunque el uso natural del lenguaje nos lleva a expresarnos así: antes de sorprenderse por el escaso peso del objeto uno no tiene (necesariamente) una creencia explícita acerca de su peso (en el ejemplo que estamos considerando el sujeto no habría reparado en tales pormenores antes de tomar el objeto), sino que simplemente se comportaba como de costumbre. En el ejemplo de Searle se trata de un vaso de plástico que uno toma mientras come en un restaurante. Podría decirse que la sorpresa fue causada por el hecho de que el sujeto tenía la creencia de que el vaso era de cristal y no de plástico y que por tanto pensaba que sería más pesado. Tenderíamos al menos a explicar así la situación, pero Searle propone que no existen esas creencias ni esos pensamientos en semejante circunstancia: el sujeto simplemente actúa.

Stroud, uno de los críticos de la primera formulación de la noción de Trasfondo en Searle, ha escrito lo siguiente en relación con las dificultades descriptivas que Searle presenta como ínsitas en el propio lenguaje:

> [Searle] admits that "there is a real difficulty in finding ordinary language terms to describe the Background" (Intentionality, p.156), but there seems to be no technical term that can be coherently introduced to describe it either. "Practices", "capacities" and "habits" are not right, because they must somehow be understood as "explicitly mental" phenomena. "Assumptions", or "presuppositions" are not right, because they imply that there are propositional contents which we entertain in some psychological mode. What is in the "Background" must be thought of only as "preintentional" or "non-representational", in other words, as "assumptions" or

> "presuppositions" that are not really assumptions or presuppositions at all (Stroud, 1991: 251).

Antes de la reformulación de la hipótesis del Trasfondo propuesta en *The Rediscovery of the Mind*, Barry Stroud había criticado, en el artículo citado, la primera formulación incidiendo en que con ella Searle pretendía explicar lo intencional en términos no intencionales y mostrar así que la intencionalidad no es primitiva o irreductible (proyecto que consideraba abocado al fracaso). Creemos, por nuestra parte, que lo hasta aquí expuesto sobre la filosofía de la mente de Searle es suficiente para constatar que el modo en que la conciencia es presentada por Searle como irreductible y el modo en que entiende que la intencionalidad necesita de la conciencia fenoménica para poder ser tenida por verdadera intencionalidad apuntan a un planteamiento antirreduccionista que, indiquemos sólo de pasada, a diferencia de Stroud no entendemos como uno de los puntos fuertes de la concepción de lo mental de Searle pues, como vimos en el capítulo anterior, conduce a un peculiar chauvinismo fenoménico a través de cauces inseparatistas. Este chauvinismo inseparatista se hace totalmente explícito con la formulación del principio de conexión y la concomitante reformulación de la hipótesis del Trasfondo en Searle (1992). Veamos cómo queda el mapa de lo mental que Searle traza con ésta reformulación de la hipótesis del Trasfondo.

En el capítulo octavo de *The Rediscovery of the Mind* Searle ofrece una nueva versión de su hipótesis del Trasfondo, de la cual dice (Searle, 1992: 175 del original, 181 de la traducción) que significa un avance respecto de la ofrecida una década antes en *Intentionality* dado que ha sido reelaborada en algunos relevantes respectos. Y es cierto: la reelaboración es significativa. La distinción entre la Red y el Trasfondo presentada previamente desaparece en esta nueva aproximación. Searle entiende que con esto no sólo derriba una frontera que trazó injustificadamente, sino que, además, con esta nueva aproximación estaría abandonando toda una concepción de la mente (en realidad, de la Red):

> On the view of the mind as containing an inventory of mental states, there must be a category mistake in trying to draw a line between the Network and the Background, because Background consists of a set of capacities, and the Network is not a matter of capacities at all, but of intentional states (…) I now think the real mistake was to suppose that there is an inventory of mental states, some conscious, some unconscious (…) The thesis of the Background has to be rewritten to get rid of the presupposition of the mind as a collection, an inventory, of mental phenom-

> ena, because the only occurent reality of the mental as mental is consciousness (Ibíd.: 187 del original, 192 de la traducción).

La centralidad de la mente consciente en Searle es la que, en último término, le lleva a rechazar su anterior concepción, que denomina ahora concepción de la mente como inventario de estados mentales. Toda nuestra vida mental consiste, según Searle, en estados conscientes y en todos aquellos estados y procesos que son capaces de generar estados conscientes. Sólo hay en el cerebro neurofisiología y conciencia, y la intencionalidad de aquellos estados que conforman la Red depende de esta segunda. Mientras un estado no es consciente, su intencionalidad se debe, en el nuevo planteamiento de Searle, a su potencialidad consciente, es decir, se interpreta en términos de la capacidad del cerebro para causar intencionalidad consciente, cosa que apenas se diferencia de lo que Searle había dicho anteriormente del Trasfondo (pero no de la Red). La distinción entre una Red intencional y un Trasfondo preintencional desaparece. Ambas cosas son ahora, en la versión revisada de Searle (1992), prácticamente la misma: capacidades del cerebro para generar intencionalidad consciente la una y capacidades del cerebro para fijar la aplicación de estados conscientes el otro.

> What goes on in the brain, other than consciousness, has an occurent reality that is neurophysiological rather than psychological. When we speak of unconscious states, we are speaking of the capacities of the brain to generate consciousness. Furthermore, some capacities of the brain do not generate consciousness, but rather function to fix the application of the conscious states. They enable me to walk, run, speak, etc. (Ibíd.:188 del original, 193 de la traducción).

Los estados intencionales inconscientes que otrora integraran la Red devienen en la nueva aproximación de Searle "capacidades del cerebro para generar conciencia". En esta nueva aproximación Searle adopta una forma de inseparatismo chauvinista que puede resumirse (caricaturizarse) como sigue: todo es neurofisiología y punto; pero sobre algunos pedazos de tejido neuronal se han derramado algunas gotitas fenoménicas con arreglo a las cuales cabe decir de ellos que portan verdadera intencionalidad. La nueva concepción del Trasfondo supone un cambio significativo en la concepción searleana del inconsciente. Dejemos que sea él quien describa las consecuencias de este cambio de enfoque para su anterior distinción entre la Red y el Trasfondo:

> Given this picture, how do we account for all those intuitions that led us to the original thesis of the Background and to the distinction between Background and Network? According to [my new] account (…), when we describe a man as having an unconscious belief, we are describing an occurrent neurophysiology in terms of

> its dispositional capacity to cause conscious thoughts and behaviour. But if that is right, then it seems to follow that the Network of unconscious intentionality is part of the Background. *The occurrent ontology of those parts of the Network that are unconscious is that of a neurophysiological capacity, but the Background consists entirely of such capacities* (…) The question of how to distinguish between Network and Background disappears, because the Network is that part of the Background that we describe in terms of its capacity to cause conscious intentionality. But we are still not out of the morass, because we are left with the question: What is to become of the claim that intentionality functions against a set of nonintentional capacities? Why is the capacity to generate the belief that Bush is president to be treated any differently from the capacity to generate the belief that objects are solid, for example? And are we to make the distinction between the functioning of unconscious intentionality and nonintentional capacities? It seems we have traded the problem of distinguishing between Network and Background for the problem of distinguishing the intentional from the non-intentional within Background capacities" (Ibíd.: 188-189 del original, 193-194 de la traducción).[52]

La nueva formulación de la hipótesis del Trasfondo puede esquematizarse atendiendo al contraste entre los cinco puntos iniciales en que Searle resume su hipótesis original (vid. Ibíd.: 177 del original, 182-183 de la traducción) y los cuatro en que resume su hipótesis revisada (Ibíd.: 190-191 del original, 196 de la traducción):

1._ (Versión original) Los estados intencionales no funcionan autónomamente. No determinan *aisladamente* las condiciones de satisfacción.

1._ (Versión revisada) Los estados intencionales no funcionan autónomamente. No determinan sus condiciones de satisfacción *independientemente*.

Los estados intencionales que otrora aparecieran en la filosofía de la mente de Searle como formando parte de una Red de intencionalidad, y por tanto determinando conjuntamente sus condiciones de satisfacción en su mutua remisión, aparecen ahora, "de hecho, aislados" (Hermoso & Chacón, 2000: 176), y puede entenderse que no funcionan autónomamente ni determinan de forma independiente sus condiciones de satisfacción porque lo hacen relativamente a un Trasfondo preintencional de estados neurofisiológicos que no son en sí mismos estados intencionales pero que mantienen el rasgo del contorno de aspecto en virtud de su capacidad para causar estados intencionales conscientes.

2._ (Versión original) Cada estado intencional requiere para su funcionamiento *una Red de otros estados intencionales*. *Las condiciones de satisfacción se determinan sólo de manera relativa a la Red*.

[52] Las cursivas son nuestras.

2._ (Versión revisada) Cada estado intencional requiere para su funcionamiento *un conjunto de capacidades de Trasfondo*. *Las condiciones de satisfacción se determinan sólo relativamente a esas capacidades*.

3._ (Versión original) Incluso la Red no es suficiente. La Red sólo funciona de manera relativa a un conjunto de capacidades de Trasfondo.

3._ (Versión revisada) Entre esas capacidades estarán algunas que son capaces de generar otros estados conscientes. A *estos otros*[53] se aplican las condiciones 1 y 2.

La Red no aparece ya alzándose sobre el Trasfondo, claramente separada y diferenciada de éste, relativamente al cual funcionaba en la versión original, sino que ahora, en la versión revisada, todo lo que puede decirse es que algunas capacidades de Trasfondo pueden generar estados mentales conscientes.

4._ (Versión original) Esas capacidades no son y no pueden ser tratadas como *más*[54] estados intencionales o como parte del contenido de algún estado intencional particular.

Este cuarto punto desaparece en la versión revisada. Entendemos que ellos se debe a la reformulación de las relaciones entre Red y Trasfondo que se deriva de los puntos 2 y 3: el Trasfondo ha absorbido ahora las funciones de la Red, y hablar explícitamente de un Trasfondo intencional resulta tan polémico como eliminar la alusión a la intencionalidad al tratar del mismo (lo primero en vista de la circularidad argumentativa, porque el trasfondo es propuesto como fundamento de la intencionalidad de unos estados intencionales incapaces de autointerpretarse y autoaplicarse, y lo segundo, precisamente, por el hecho de que en la versión revisada el trasfondo absorbe las funciones de la red). Así, a fin de preservar la posibilidad de intencionalidades inconscientes, Searle conservará el contorno de aspecto para el Trasfondo en virtud de la virtualidad o potencialidad consciente que tratará de asegurar mediante el principio de conexión. En este sentido, la formulación revisada no hará referencia explícita al Trasfondo como instancia no-representacional: al principio la Red funcionaba contra un Trasfondo no-intencional y no-representacional, ahora, en la versión definitiva, el Trasfondo hace funcionar toda intencionalidad y toda representación, del mismo modo que de él depende toda fenomenalidad –dado que la propia intencionalidad, como defenderá Searle mediante su princi-

[53] La traducción española dice en este punto "estas otras", cuando (como han sabido ver Hermoso & Chacón, 2000) es obvio que las condiciones 1 y 2 se aplican no a las capacidades de Trasfondo sino a los estados conscientes.

[54] La traducción española (pág. 183) dice en este punto «tratadas como *meros* estados intencionales», cuando el original inglés (pág. 177) dice «treated as *more* intentional states» (las cursivas son nuestras). Esta errata, hasta donde sabemos, se le ha escapado a todos los comentaristas de habla hipana.

pio de conexión, depende del carácter fenoménico de la experiencia consciente–. Vemos pues que Searle pone no poco peso a hombros de su Atlas, el Trasfondo (resulta natural, si se nos permite el chiste, que escriba siempre «Trasfondo» con mayúscula).

5._ (Versión original) El mismo contenido intencional puede determinar diferentes condiciones de satisfacción (tales como las condiciones de verdad) y con relación a algún Trasfondo no determina ninguna en absoluto.

5._ (Versión revisada) El mismo *tipo* de contenido intencional puede determinar diferentes condiciones de satisfacción cuando se manifiesta en diferentes instancias conscientes, de manera relativa a diferentes capacidades de Trasfondo, y relativamente a algunos Trasfondos no determina ninguna.

Comencemos por sumar lo contenido sobre el segundo Trasfondo de Searle en el contraste entre aquellos cinco puntos iniciales y estos cuatro definitivos a lo que ya sabemos sobre su postura acerca del problema de la conciencia para poder atender brevemente así al paisaje encerrado en una esquemática panorámica en la que la reformulada hipótesis del Trasfondo aparece ocupando su lugar dentro del cuadro general de la filosofía de la mente de Searle. Es claro que el principal cambio de perspectiva tiene que ver con la inclusión de la conciencia como ingrediente de su concepción del inconsciente. Searle considera que su anterior punto de vista sobre el Trasfondo estaba lastrado por su concepción de la mente como inventario de estados mentales, conscientes unos e inconscientes otros en un determinado momento, pero igualmente mentales e intencionales ambos en todo momento. Esa perspectiva, entiende, no hacía justicia a la conciencia como fenómeno mental crucial: "la conciencia no era esencial a los fenómenos mentales" (Searle, 1992: 186 del original, 192 de la traducción). El cerebro, nos dice ahora, alberga una enorme cantidad de neuronas incrustadas en células gliales, y algunas veces la conducta de los elementos de esta compleja masa mucilaginosa causa estados conscientes. En determinadas circunstancias, por otra parte, estados particulares del sistema son tales que *podrían* causar estados conscientes, y entonces, sólo entonces, nos hallamos ante intencionalidad (y, por ende, *mentalidad*) inconsciente. Lo que sucede en el cerebro y podemos designar como mental, pero que al tiempo no es consciente en un momento dado, no son sino capacidades del cerebro para generar conciencia (esta idea es la que trata de sustentar el principio de conexión). Además, algunas capacidades del cerebro no generan conciencia, sino que más bien funcionan para fijar la aplicación de los estados conscientes: capacitan para andar, escribir, hablar… En la formulación defi-

nitiva (Searle, 1992) de la hipótesis del Trasfondo, las relaciones entre ese Trasfondo capacitador y la Red de intencionalidad aparecen transformadas –respecto de la primera formulación de la hipótesis del Trasfondo en Searle (1983)–. Searle (1992: 188 del original, 194 de la traducción) entiende que al describir una creencia inconsciente describimos una neurofisiología ocurrente en términos de su capacidad disposicional para causar eventos conscientes. Pero si esto así, entonces la Red de intencionalidad inconsciente es parte del Trasfondo, pues la ontología ocurrente de aquellas partes de la Red que son inconscientes es la de una capacidad neurofisiológica, y el Trasfondo consta enteramente de tales capacidades. Así, la cuestión de cómo distinguir entre Red y Trasfondo desaparece porque la Red es aquella parte del Trasfondo que describimos en términos de su capacidad para causar intencionalidad consciente. El problema sería ahora el de distinguir lo intencional de lo no intencional dentro de las propias capacidades de Trasfondo. Searle, de un modo un tanto tentativo, responde a dicho problema mediante las siguientes distinciones:

1._ Necesitamos distinguir entre lo que está en el centro de nuestra atención consciente de las condiciones límite, periféricas y de situación de nuestras experiencias conscientes.

2._ Necesitamos distinguir dentro de los fenómenos mentales la forma representacional de la no representacional. Puesto que la intencionalidad se define en términos de representación, ¿cuál es el papel, si es que hay alguno, de lo no representacional en el funcionamiento de la intencionalidad?

3._ Necesitamos distinguir capacidades de sus manifestaciones. ¿Y cuáles de las capacidades del cerebro deberían pensarse como capacidades de Trasfondo?

4._ Necesitamos distinguir aquello en lo que nos interesamos efectivamente de aquello que damos por sentado.

Mas, contemplando estas distinciones, cabe dudar que Searle solucione con ellas el referido problema, dado que mientras la primera y la última apenas nos dicen nada acerca de su posible aplicación de cara a ofrecer la solución requerida, las dos centrales consisten en poco –si acaso algo– más allá de matizadas reformulaciones del propio problema que tratan de contribuir a resolver.

Una vez abandonada la concepción de la mente como inventario, la remozada hipótesis del Trasfondo vendría a consistir en lo siguiente: "toda la intencionalidad consciente –todo pensamiento, percepción, comprensión, etc.– determina condiciones de satisfacción sólo relativamente a un conjunto de capacidades que no son y no pueden ser parte de ese mismo estado consciente. El contenido efectivo por sí mismo es insuficiente para determinar las condiciones de satisfacción" (Ibíd.: 189 del original, 195 de la traducción). Esta hipótesis, según Searle (1999c), ofrece una aproximación más parsimoniosa y cabal a las relaciones entre la mente consciente, la intencionalidad y el cerebro que la habitual dentro del paradigma cognitivo. "En lugar de pensar en el cerebro como lo que lleva a cabo una grandísima cantidad de operaciones de acuerdo con reglas, deberíamos pensar en los procesos inconscientes del cerebro como algo que forma un conjunto de capacidades que hacen posible que nuestra vida consciente funcione, y esas capacidades son lo que yo llamo el Trasfondo. El Trasfondo es un conjunto de capacidades habilitadoras" (Searle, 1999c: 9τ). Así, de la hipótesis original del Trasfondo Searle conserva la idea de que incluso después de hacer explícitos todos los contenidos de la mente como un conjunto de reglas, pensamientos, creencias, etc. conscientes, aún se requiere de un conjunto de capacidades de Trasfondo para su interpretación, mientras que desecha la idea de la realidad ocurrente de la Red inconsciente de intencionalidad (es decir, la idea una Red de intencionalidad claramente separada del Trasfondo, la idea de una mínima posibilidad de existencia para estados intencionales allí donde no haya experiencia consciente: a esto se refiere con "concepción de la mente como inventario"). Por su parte, la distinción entre lo intencional y lo no-intencional dentro del Trasfondo, pretende Searle, por más que de forma un tanto insatisfactoria, se diluiría por cuanto toda vez que comencemos a pensar conscientemente en cualquiera de las capacidades, habilidades o presuposiciones que integran el Trasfondo, estaríamos en disposición de formarnos creencias –conscientes– sobre ellas, de modo que las mismas pasarían a contar como creencias iguales a cualquier otra–salvo que, talvez, mucho más generales.

Con todo, es evidente que residuos de la distinción original entre Red y Trasfondo persisten y se muestran reluctantes a disiparse con la hipótesis reformulada del Trasfondo, porque, además de esas capacidades de Trasfondo que efectivamente generan estados intencionales conscientes –que se presentan como residuos de la Red de la versión original, residuos que se hacen explícitos en el tercer punto de la hipótesis revisada: "entre esas capacidades estarán algunas que son capaces de generar otros estados conscientes"–, Searle, como hemos visto, nos habla de *otras* que, "sencillamente" (entreco-

millamos porque precisamente con estas *otras* estamos de vuelta en las dificultades que la reformulación trataba de solventar), contribuyen a la aplicación de estados intencionales conscientes y capacitan para hacer cosas tales como andar, comer, escribir o hablar, y *estas otras capacidades de Trasfondo* –residuos del Trasfondo tal y como fuera concebido originalmente– son las que resulta ahora difícil distinguir de aquéllas directamente relacionadas con la habilitación causal de estados intencionales conscientes que el Trasfondo searleano pretende ofrecer; y sin *estas otras capacidades de Trasfondo* Searle se vería abocado a la circularidad argumentativa (dado que su búsqueda de una base preintencional para la aplicación de estados intencionales incapaces de autointerpretarse habría recurrido, en el fondo, a otros estados intencionales, que aparecen en la versión revisada de su hipótesis como esa parte del Trasfondo que hemos caracterizado como un residuo de la Red de la hipótesis original) y se mostraría incapaz de hacer frente a la deseada conceptualización anticognitivista de los problemas que le llevaron a la formulación original de su hipótesis del Trasfondo (entre los cuales destacaban la comprensión del significado y la adquisición y ejecución de habilidades y, decisivamente, el que los estados intencionales que conformaban la Red original no se autointerpretaran, de donde surgía la necesidad del Trasfondo, un Trasfondo problemático en la primera formulación por sus difusos límites externos, esto es, sus límites con la Red, e igualmente problemático en la última dadas unas demarcaciones internas igualmente difusas: el problema de la autointerpretación se resolvía en la primera formulación distinguiendo entre lo representacional y lo pre-representacional, pero esa distinción que no era clara entonces no llegó a serlo después mediante su difuminación en el parcheado al que Searle sometió su esquema teórico). La claridad de esa frontera entre lo intencional y lo no intencional en el remozado Trasfondo no es, pues, una cuestión baladí, pero tampoco una cuestión que el esquema teórico de Searle se muestre capaz de resolver.

El mapa definitivo del Trasfondo consta de estados neurofisiológicos caracterizados como capacidades para causar determinados estados intencionales conscientes (la antigua Red), y estados neurofisiológicos caracterizados como capacidades que hacen posible toda forma de intencionalidad (el antiguo Trasfondo). El problema que Searle advirtiera en su primera formulación y le llevara a la segunda aparece idéntico en el nuevo mapa, pero bajo denominaciones, acaso, más sofisticadas. La necesidad de fusionar Red y Trasfondo surge para Searle con la inclusión del principio de conexión, porque con él las ontologías de ambos se presentan como una y la misma en un importante sentido (cuando los estados intencionales que conformaban la Red son inconscientes,

como veíamos, aparecen como capacidades neurofisiológicas para causar estados intencionales conscientes, pero esa ontología, la de meras capacidades neurofisiológicas, era la antes reservada para el Trasfondo, motivo por el cual Searle arrostra la señalada fusión como única posibilidad dentro de su esquema teórico). Mas, tras la fusión, como indicábamos, los problemas permanecen y no sabemos con certeza qué es lo que en el Trasfondo contribuye a la causación de estados intencionales concretos, ni cómo lo hace, ni qué es lo que en el Trasfondo contribuye a fijar la aplicación de cualesquiera estados intencionales conscientes, ni cómo lo hace, ni tampoco cómo distinguir entre ambos segmentos del Trasfondo. De este modo, también la distinción entre intencionalidad inconsciente (mentalidad inconsciente) y meros fenómenos neurofisiológicos (neurofisiología no-mental) permanece envuelta en las brumas de la indeterminación en la nueva formulación. En resumidas cuentas, en un movimiento que genera, esencialmente, los mismos problemas que venía a solucionar, las fronteras difusas entre Red y Trasfondo son sustituidas por unas evanescentes demarcaciones internas entre lo intencional y lo no intencional dentro del propio Trasfondo y por unos límites indefinidos entre lo mental y lo neurofisiológico (vid. Hermoso & Chacón, 2000: 186).

Tras la versión revisada de la hipótesis del Trasfondo de 1992, Searle incidiría, en un artículo publicado siete años más tarde, en algunos de los extremos decisivos allí tratados, pero apenas sumaría nada a lo dicho. En dicho artículo (el último de sus escritos sobre el Trasfondo), pues, Searle (1999c: 9) propone una vez más que existe un argumento convincente a favor de la existencia del Trasfondo, y es, nuevamente, el de que los contenidos intencionales son por sí mismos insuficientes para fijar sus condiciones de satisfacción, cosa que sólo pueden hacer, dice, en relación a un Trasfondo de capacidades y habilidades –no de *reglas*, lo cual resulta crucial para su argumentación– que no son ellas mismas parte de los contenidos intencionales. Así, según Searle, por ejemplo, nuestro acceso a la semántica se apoya en un suelo de capacidades y habilidades que nada tienen que ver con la semántica, sino más bien con nuestra filiación y nuestra fijación bio-histórica, con nuestro histórica y biológicamente radicado modo de habitar nuestro espacio-tiempo.

El Trasfondo, arguye Searle (1999c: 9) nuevamente (al igual que en Searle, 1983 y en Searle, 1992), funciona causalmente, pero, insiste el de Denver, es el suyo un tipo particular de causalidad al que nuevamente (tras dieciséis años) llama habilitador y sobre el que nuevamente poco dice explícitamente más allá de que la forma de la causali-

dad que opera a través del Trasfondo es la forma de la "habilitación", con lo cual quiere decir que el Trasfondo concurre y coadyuva a hacer posibles gran cantidad de virtualidades de nuestra mente, y que pese a que el Trasfondo es una capacidad causal no hay en él condiciones causalmente suficientes (1999c: 13), sino sólo condiciones habilitadoras para que pueda operar determinado tipo de intencionalidad (Ibíd.: 13-14). Como puede verse, tras dieciséis años, al abundar en este extremo de la causalidad del Trasfondo se limita Searle a la repetición literal de lo dicho: no amplía, aclara o discute sus anteriores –y escasas– afirmaciones acerca del funcionamiento causal del Trasfondo. Searle incide en que nuestra aptitud para la interpretación lingüística se apoya en capacidades de Trasfondo, y aduce que dadas diferentes capacidades de Trasfondo nuestra interpretación sería diversa, cosa que presenta una vez más como prueba o argumento a favor del Trasfondo y su causalidad habilitadora. Además de éste nivel lingüístico, Searle (Ibíd.: 12 y ss.) hablará de otros niveles del Trasfondo, como el perceptivo, el motivacional o el relacionado con la experiencia consciente. Respecto de este último nivel, sostiene que las capacidades de Trasfondo ejercen un influjo omnipresente y crítico sobre la conciencia, dado que toda experiencia consciente (recuérdese el tercer rasgo secundario de la propuesta descriptiva que Searle denomina *estructura de la conciencia*) tiene lugar dentro de una gradación de *familiaridad*, y ésta está obvia e inextricablemente entrelazada con las capacidades de Trasfondo que, según Searle, poseemos en virtud de nuestra historia biológica –Searle (1983) hablaba, como veíamos, de un "Trasfondo profundo" en relación con la misma– y cultural –Searle (1983), como asimismo veíamos, hablaba de un "Trasfondo local" en relación con ésta.

Para acabar con este compendioso repaso del mapa definitivo de lo mental que Searle traza y defiende, enumeramos las leyes del funcionamiento del Trasfondo que postula (indicando que no pretende ser exhaustivo):

1._En general, no hay acción sin percepción y no hay percepción sin acción.

2._La intencionalidad ocurre en un flujo coordinado de acción y percepción, y el Trasfondo es la condición de posibilidad de las formas tomadas por el flujo.

3._La intencionalidad tiende a elevarse al nivel de la capacidad de Trasfondo.

4._Aunque la intencionalidad suba al nivel de la capacidad de Trasfondo, alcanza la capacidad hasta el fondo.

5._El Trasfondo sólo se manifiesta cuando hay contenido intencional.

La noción de Trasfondo y, decisivamente, el modo en que ésta fuera reformulada mediante la inclusión del principio de conexión constituye un importante flanco de la crítica de Searle al paradigma cognitivo y, concretamente, al computacionalismo, pero quizá el más destacado y asimismo comentado sea el que propusiera con su experimento mental de la habitación china.

### 4.2._La Habitación china

El experimento mental de la habitación china surgió en 1977, en Berkeley, a partir de una conferencia de Roger Schank acerca de su trabajo con Robert Abelson en la elaboración de programas computacionales para la comprensión lingüística de narraciones cuya acción tenía lugar en situaciones bien circunscritas. A pesar de que Searle criticara el planteamiento de Schank tras la intervención de éste (Dreyfus, 1979: 311), el famoso experimento mental no formó parte de la crítica de Searle durante aquella sesión, sino que sólo sería desarrollado posteriormente (Faigenbaum, 2003: 78) y no aparecería publicado hasta 1980, en un artículo para *The Behavioral and Brain Sciences* intitulado "Mind, brain, and programs". Antes de su publicación, durante los últimos años de la década de los setenta, el experimento mental fue expuesto por Searle ante diferentes audiencias, de forma que fue recibiendo algunas críticas que pasaría a incluir (de forma anónima) y responder en su artículo de 1980. Además, siguiendo la fórmula editorial de *The Behavioral and Brain Sciences*, en la publicación original del artículo de 1980, al núcleo del mismo (que ocupaba las páginas 417-424 del volumen en que apareció) seguía una sección que contenía una considerable cantidad de críticas y comentarios de diversos autores (pp. 424-450), tras la cual Searle cerraba el artículo con una breve respuesta a los mismos (pp. 450-456). Aunque Searle (1984b; 1984c; 1990d; 1997a) ha formulado su experimento mental de diferentes maneras, ciertamente, ningún cambio de excesiva entidad (aunque sí cambios de acentuación y ligeros matices) justifica las casi idénticas reformulaciones aparecidas en sucesivas publicaciones. Comentaremos pues el experimento mental de la habitación china y la argumentación paralela al mismo siguiendo el orden de la presentación que Searle hiciera de ambos en este artículo, e integraremos en nuestra exposición de los mismos tal y como aparecieran originalmente algunos matices provenientes de las mencionadas reformulaciones posteriores.

La pregunta de la que parte Searle en este comentadísimo artículo es la siguiente: ¿qué relevancia filosófica y psicológica hemos de conceder a los esfuerzos realizados con la intención de simular computacionalmente las capacidades cognitivas humanas? (Searle 1980: 183)[55]. Searle ensaya su respuesta partiendo de algunas precisiones: en primer lugar bosqueja su conocida distinción entre inteligencia artificial débil ("weak" or "cautious" AI) y fuerte ("strong" AI), y apunta que la débil, desde el punto de vista de la cual las computadoras no son sino herramientas útiles en el estudio de la mente del mismo modo que los son en otras áreas de la investigación científica, no será objeto de su crítica. La misma será dirigida, pues, a la que denomina inteligencia artificial fuerte (IA fuerte), y en particular a la tesis que atribuye a ésta según la cual una computadora adecuadamente programada tiene, literalmente, estados cognitivos –Searle escribe «understand» en cursivas en este punto, en la frase "*understand* and have other cognitive states" (1980: 183), barruntando el tipo de crítica que realizará– y, por lo tanto, los programas pertinentes explican la cognición humana o –como dice la traducción española contenida en Boden (1990: 82)– constituyen[56] por sí mismos una explicación de ésta. Searle hace una segunda precisión previa a la presentación del *Gedankenexperiment* de la habitación china y su concomitante argumentación: la de que tendrá en mente a lo largo de la misma los trabajos de los años setenta de Roger Schank y sus colaboradores (cita como referencia el libro de Schank y Abelson *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*), pero que su argumento se aplica por igual a tales trabajos, los de Joseph Weizenbaum (Searle menciona de pasada el célebre programa que éste diseñara en el MIT entre 1964 y 1966: ELIZA, que utilizaba patrones almacenados para reconocer frases sin computar su estructura sintáctica y que remedaba una consulta psicoterapéutica rogeriana), los de Terry Winograd (Searle se refiere, nuevamente sin entrar en detalles, a SHRDLU, un programa diseñado entre 1968 y 1970 que, a diferencia de ELIZA, disponía de un procesador sintáctico destinado a organizar información relativa a un mundo de figuras geométricas que constituía el entorno de dicho programa)[57] y en realidad a "cualquier simulación en máquina de Turing de fenómenos mentales humanos" (Searle 1980: 183τ). Searle ofrece una sucinta

[55] Citamos siguiendo la reimpresión contenida en Haugeland (1997).
[56] Mencionamos la traducción por el inocente matiz indirecto que confiere a la frase al verter "are" por "constituyen".
[57] Es interesante hacer notar que el influyente manual de psicolingüística de Trevor A. Harley, sin negar enteramente las capacidades semánticas de SHRDLU, dice de éste y de ELIZA que tenían capacidades de procesamiento sintáctico muy primitivas y que carecían de la *potencia de cálculo* necesaria para analizar el lenguaje humano (Harley, 2014: 15 del original, 10 de la traducción).

descripción de lo que tiene en mente cuando en su ataque al computacionalismo dice "cualquier simulación en máquina de Turing". Evitando entrar en detalles, sugiere que los programas de Schank que toma como referencia funcionan del siguiente modo (recordemos que según él la elección de los programas de Schank no tiene demasiada relevancia, dado que su argumento se aplicaría por igual a "cualquier simulación en máquina de Turing de fenómenos mentales humanos", y a pesar de esto, su lacónica descripción evidencia una atención unilateral a programas basados en guiones o scripts): dada una historia determinada (en su ejemplo, la de un hombre que va a un restaurante y sale del mismo descontento y sin pagar) y un marco conceptual (el de los restaurantes, en su ejemplo), pueden dichos programas responder a preguntas acerca de la historia. Según Searle, la IA fuerte defendería que los programas *indicados*[58] entenderían las historias acerca de las cuales responden preguntas y serían de hecho explicaciones de la habilidad humana de hacer tales cosas. Así las cosas, Searle tiene en mente los primeros intentos de simular el lenguaje humano, es decir, programas capaces de ofrecer respuestas del tipo de las necesarias para superar el test de Turing: una prueba (en realidad un "juego") que implica, como el *Gedankenexperiment* de Searle, una habitación de la que salen tipografiadas una serie de respuestas a preguntas que son formuladas a lo que quiera que haya dentro de la misma. Veamos en qué consiste dicho test en un pequeño excurso antes de explicar el experimento mental que Searle propusiera en su artículo de 1980.

El famoso test de Turing fue planteado en un artículo publicado en *Mind* en 1950 bajo el título "Computing machinery and intelligence"[59] por Alan Mathison Turing, a la sazón director adjunto del laboratorio de computación de la Universidad de Manchester. Este artículo plantea un interrogante que en ocasiones ha sustituido al título original en las traducciones españolas: ¿Puede pensar una máquina? La pregunta sería clara… si pudiéramos definir los términos que la integran. Pero, ¿cómo? Turing plantea al comienzo del artículo la señalada pregunta y propone que resulta inabordable si lo que buscamos al formularla es ofrecer una definición de los términos «pensar» y «máquina» partiendo del uso ordinario de los mismos. Turing ofrece esta perspectiva en un momen-

[58] Las cursivas quieren apuntar aquí que una cuestión decisiva desde el punto de vista de la IA fuerte, aunque no desde la perspectiva de Searle, sería la siguiente: ¿es la clase de programas indicada por Searle la clase *indicada* de programas de cara al planteamiento de los extremos que Searle pone sobre la mesa?

[59] La evidente relación entre estos dos artículos (vid., v. g., Copeland, 1993: 121-139; Crane, 1995/2003: 123; Martínez-Freire, 2007a: 167; Vilarroya, 2002b: 101 y ss., especialmente 103) ha llevado a algunos compiladores de antologías de textos fundamentales en ciencias cognitivas y filosofía de la mente a situarlos uno tras otro en sus antologías (así, v. g., en Cummins & Cummins, 2000).

to en que se impone en la filosofía inglesa el análisis del lenguaje cotidiano (la obra de referencia de Ryle *The Concepto of Mind* fue publicada sólo un año antes de la aparición del artículo que nos ocupa), lo cual hace pensar en una actitud de reserva por su parte hacia tal tendencia cuando de lo que se trata es de resolver problemas a medio camino entre lo conceptual y lo empírico. Una definición de este tipo, esto es, ensayada desde la consideración del uso ordinario de los términos mentados, es considerada por Turing absurda por cuanto conduciría a algo así como un compromiso de carácter estadístico con los resultados de una encuesta sociológica. En lugar de esa vía, pues, opta el lógico inglés por abordar la cuestión diseñando el famoso juego de la imitación o test de Turing. Con este juego Turing sustituye una pregunta que contiene términos abstrusos por otra que considera más clara: ¿Puede una máquina engañar a una persona con la misma eficacia que otro ser humano? El juego de la imitación está diseñado para ofrecer la posibilidad de dar respuesta a esta pregunta. En el mismo participan tres personas: un hombre, una mujer y un interrogador que debe adivinar el sexo de ambos mientras uno de los dos trata de confundirle o engañarle con sus respuestas. El interrogador no puede ver ni oír a los participantes, de modo que en sus indagaciones ha de atenerse exclusivamente a las respuestas ofrecidas por ambos participantes a las preguntas que libremente formula. Parece que el interrogador sólo cuenta pues con la posibilidad de evaluar aspectos mentales de las respuestas de los participantes a la hora de determinar quién es el hombre y quién la mujer, pero, de hecho, lo que evalúa no es sino la *conducta* verbal de ambos (que, a fin de evitar otro tipo de pistas, le llega mecanografiada). En estas circunstancias, Turing propone que una máquina sustituya al participante que pretende engañar al interrogador en su labor de identificación, y que si logra en efecto engañarle habremos alcanzado un punto en el que la respuesta a la pregunta de partida se inclinaría hacia el sí: la máquina puede pensar.

Turing ofrece réplica a lo largo del artículo a diversas objeciones que, considera, pueden esgrimirse contra su planteamiento. De entre las mismas es interesante destacar aquí una que Turing despacha rápidamente como irrelevante. Se trataría de valorar si no cabe la posibilidad de que entre las facultades mentales humanas y las de una supuesta máquina inteligente no habría un importante hiato; esto es, se trataría de valorar si no podríamos encontrarnos ante procesos radicalmente diferentes a pesar de las indiscernibles respuestas conductuales. Turing concluye que frente a la prueba que él propone la objeción carece de objeto, pues, según su planteamiento, superar la prueba es acceder a un mismo mundo, el de las *cosas* que piensan.

De vuelta ya de nuestra breve digresión, pasemos a describir el popular experimento mental de la habitación china, que se presentara originalmente como una refutación tanto de la validez del test de Turing como criterio de mentalidad como del programa simbolista clásico en inteligencia artificial (Bermúdez, 2010/2014: 163), una refutación que Searle terminaría presentando como extensible a toda forma de computacionalismo en ciencias cognitivas. Searle (1980), como apuntábamos, tendrá en mente los programas de Schank a la hora de plantear su *Gedankenexperiment*. Los mismos, como apuntábamos, fueron diseñados para ofrecer respuestas a preguntas concretas acerca de historias breves que les eran proporcionadas previamente (dichas preguntas podían referirse a aspectos de las historias no explícitos en el texto de las mismas). A tal efecto, los programadores dotaron a su software de un tipo especial de base de datos (denominada base de conocimiento) acerca de características típicas del contexto de las historias. Responder a las preguntas significaría para este tipo de software detectar características relevantes que aparezcan a la vez en la pregunta y en el texto de la historia y deducir, usando su base de datos, extremos por los que se le pregunta y que no aparecen específicamente en el texto de la historia. El texto de la historia podría, por ejemplo, describir un funeral. Ante una pregunta como: "¿Condujo la hija del finado charlando y riendo hacia el tanatorio?", cuya respuesta no se halla explícitamente en el texto de la historia, el software usaría su base de datos para enlazar aspectos que van unidos característicamente entre sí en situaciones como la descrita por la historia y respondería algo así como: "Muy probablemente la hija del finado no condujera riendo y charlando hacia el tanatorio". O podría describir un parto; en este caso, ante la pregunta "¿fue cortés el alumbrado al saludar a los asistentes?", el programa respondería que no, ya que su base de datos incluiría referencias a la nula capacidad lingüística de los neonatos.[60]

Con este tipo de software en mente, Searle (1980: 184 y ss.) nos pide que nos lo imaginemos encerrado en una habitación dentro de la cual recibe un considerable legajo (a large batch) de escritos en lengua china. Searle, encerrado en su habitación, desconoce dicho idioma y es de hecho incapaz de distinguir la escritura china de la japonesa o de dibujos sin sentido. Nos pide asimismo que imaginemos que tras ese primer lote de

[60] Para más detalles, consultar, por ejemplo, la descripción de SAM (el primer software para la comprensión de historias que Schank desarrollara en Yale) que Cullingford ofrece en Schank & Riesbeck (1981: 75-132), o las alusiones del propio Schank al mismo en el primer capítulo de Schank (1999: especialmente pp. 5 y ss.), o (para una sucinta perspectiva de las relaciones entre el aludido programa y otros semejantes) el séptimo capítulo de Rose (1985) o, particularmente, Schank & Abelson (1977) y Schank & Yale A. I. Project (1975).

escritos en chino le es entregado un segundo lote de escritos (nuevamente pictogramas, ideogramas y fonogramas chinos) junto con un conjunto de reglas escritas en inglés (idioma que, claro, entiende a la perfección) para relacionar (fragmentos de, hemos de entender: Searle (Ibíd. 185) sólo dice "rules for correlating the second batch with the first batch") ambos lotes de escritura china. Dichas reglas le permiten correlacionar un conjunto de símbolos formales con otro conjunto de símbolos formales, y lo único que pretende significar aquí con "formal" es que puede identificar los símbolos por sus formas. Nos pide a continuación que supongamos que se le entrega una tercera tanda de símbolos chinos junto con algunas instrucciones, nuevamente en inglés, que le permiten correlacionar elementos de esta tercera tanda de símbolos chinos con los símbolos de los dos lotes anteriores, instrucciones que le indicarían cómo ofrecer determinados símbolos chinos en respuesta a los que le son entregados en la tercera tanda. En este punto nos dice Searle que, sin él saberlo, quienes desde fuera de la habitación le entregaron los textos chinos llaman al primer lote "guión" (la aludida base de datos especial del modelo de Schank, la base de conocimiento), al segundo "historia", al tercero "preguntas", a los símbolos que él entrega en respuesta al tercer lote "respuestas a las preguntas" y al conjunto de reglas en inglés "programa".[61] De este modo, al ofrecer dichas "respuestas" siguiendo las reglas escritas en ingles "Searle, supuestamente, se comporta exactamente igual que una computadora ejecutando el programa: realiza operaciones computacionales sobre elementos especificados formalmente y produce sus respuestas manipulando símbolos formales no interpretados" (Moural, 2003: 217τ).

Al adquirir pericia siguiendo las instrucciones para manipular símbolos cuyo significado desconoce, propone Searle, desde el punto de vista de alguien que se encuentre fuera de la habitación, sus "respuestas" a las "preguntas" son indistinguibles de las que darían hablantes nativos de chino (Searle habría pasado el test de Turing). Pero Searle afirma que él sólo produce tales respuestas manipulando símbolos formales no interpretados y se comporta pues como un computador: realizando operaciones computacionales sobre elementos formalmente especificados. Por lo que al chino se refiere, Searle dice que dentro de su habitación él no es más que una realización de un programa

[61] A pesar de que en esta primera versión publicada de su experimento mental Searle se mantiene muy pegado a los modelos computacionales de Schank, en versiones posteriores ha simplificado su exposición reduciendo el número de lotes de texto que le son entregados (vid. Searle 1984b: 32; 1990d: 20). La idea que en cualquier caso mantiene es la de seguir una serie de reglas cuyo significado comprende para ofrecer como output símbolos cuyo significado desconoce en respuesta a insumos de símbolos cuyo significado igualmente desconoce.

computacional que habría pasado el test de Turing sin entender una sola palabra de ese idioma.

Tras exponer su experimento mental Searle vuelve sobre la aludida doble tesis de la IA fuerte: a) una computadora adecuadamente programada tiene, literalmente, estados cognitivos, entre los cuales Searle subrayaba al principio del artículo la comprensión –ahora, tras proponer su *Gedankenexperiment*, no la subraya sino que se refiere exclusivamente a ella (Searle, 1980: 185-186)–; y b) los programas pertinentes explican la cognición humana. La primera parte dicha doble tesis cae según Searle porque él –que ha pasado de un párrafo de su artículo al siguiente de ser una *realización de un programa informático* ("an instantiation of [a] computer program") a ser la propia computadora–, sentado en su habitación, no entiende nada a pesar de que sus respuestas resulten desde fuera de la habitación indistinguibles de las que daría un hablante chino perfectamente competente. Según Searle, las computadoras y sus programas no son ni más ni menos que lo que él es sentado en su habitación. Dado que en él no se da comprensión de los símbolos que manipula, tampoco se dará en ellas, porque "ninguna computadora digital, sólo en virtud de ejecutar un programa, tiene nada que [Searle sentado en la habitación china] no tenga" (Searle, 1984b: 33τ). La segunda caería porque los programas informáticos no nos aportan ni condiciones suficientes ni necesarias para la comprensión. Según Searle, su experimento mental hace evidente que las computadoras y sus programas no nos ofrecen condiciones suficientes para la comprensión por cuanto su ejemplo debiera hacer evidente la idea de una computadora en funcionamiento en la que no hay comprensión en ninguna parte (excepto por la que desde fuera puedan atribuirle). No obstante, Searle se pregunta si las computadoras y sus programas ofrecen alguna condición necesaria o contribuyen significativamente a la comprensión. Searle responde señalando que uno de los postulados de la IA fuerte consiste en la defensa de una tesis según la cual cuando una persona comprende una historia en su idioma, lo que hace es exactamente lo mismo –o quizás más de lo mismo– que lo que hace él en su habitación al manipular símbolos chinos no interpretados. Searle dice que su experimento mental no demuestra que lo que se hace necesario para la comprensión sea, sencillamente, más manipulación "formal" de símbolos, pero apunta que resultaría increíble en vista del mismo: su *Gedankenexperiment* sugiere, pero no demuestra, como Searle admite, que los programas computacionales son irrelevantes para la explicación de la capacidad humana para la comprensión de historias. En el caso del chino, Searle sentado en su habitación posee, nos dice, todo lo que la IA podría asignarle en términos de programas

computacionales, y aun así no entiende nada. El inglés sigue entendiéndolo a la perfección, y tiene la impresión de que no se ha ofrecido hasta el momento ninguna razón que haga suponer que su comprensión de su lengua vernácula tenga algo que ver con programas computacionales, esto es, en sus términos, con operaciones computacionales sobre elementos especificados formalmente. Resumiendo, Searle cree que de su experimento mental es dable colegir que los programas computacionales no ofrecen condiciones suficientes para la comprensión, y que carecemos de razones para suponer que constituyan condiciones necesarias para la misma o, siquiera, que su contribución a la explicación de la comprensión pueda ser significativa.

Searle considera que la fuerza de su argumento no radica en el hecho de que máquinas diferentes puedan tener idénticos inputs y outputs operando bajo diferentes "principios formales" (¿sigue usando "formal" en el mismo sentido?; Searle no lo aclara), sino en que cualquier "principio formal" será insuficiente para dar cuenta del entendimiento: después de todo, Searle, sentado en su habitación y siguiendo tan bien como quepa imaginar "principios formales" tan buenos como quepa imaginar no entiende nada en absoluto. Además, insiste Searle, carecemos de razones para suponer que tales "principios formales" sean necesarios o siquiera que contribuyan a la comprensión, ya que no se han ofrecido razones para suponer que cuando un hablante dado comprende frases en su idioma "opera bajo algún programa formal" (Searle, 1980: 187τ) – ciertamente, la noción de lo formal tal y como es definida inicialmente por Searle, no parece adecuarse demasiado bien a lo que parece requerir de ella en el contexto de esta cita.

Hemos seguido paso a paso, casi literalmente, a Searle en la formulación de su famoso experimento mental. Veamos ahora brevemente algunas de las ideas destacables que incluye la segunda parte de su artículo, dedicada a responder las críticas recibidas por su argumento antes de su publicación. La réplica de los sistemas (Searle en la habitación no entiende, pero sí el conjunto, el sistema conformado por él, los códigos, los elementos de los que hace uso para realizar sus cálculos, etc.) es rechazada mediante la sugerencia de una versión alternativa del experimento: Searle memoriza los símbolos y las reglas e incorpora así el sistema completo (que pasa a constituir así una parte de Searle) pero, al ofrecer sus "respuestas" en chino, sigue sin entender nada. Y una subréplica de la réplica de los sistemas (según la cual el subsistema que Searle incorporaría al memorizar el conjunto de reglas y símbolos comprende chino aunque Searle como sistema que lo abarca no lo haga) recibe la contrarréplica siguiente: Searle, cuando entien-

de inglés, conoce el significado de las palabras que son el caso, mientras el subsistema chino sólo sabe que al símbolo no interpretado X sigue el símbolo carente de sentido Y; es decir, sólo sabe que determinados símbolos ingresan y son egresados después de una serie de manipulaciones de acuerdo con unas reglas escritas en inglés (el subsistema que Searle habría incorporado estaría en la misma situación que Searle en la habitación: pasa el test chino de Turing sin entender chino en absoluto); de hecho, propone Searle, tal subsistema siquiera sería un subsistema diferente del inglés, sino una parte suya capaz de manipular símbolos que no comprende. En su comentario a esta réplica de los sistemas Searle menciona explícitamente el test de Turing y, crucialmente, su perspectiva –que desarrollará en próximos escritos– según la cual perder de vista la distinción entre intencionalidad intrínseca e intencional metafórica conlleva una expansión absurda de las adscripciones de intencionalidad. En esta ocasión no lo expresa así, sino que señala que si definimos cognición como aquello que sucede cuando tenemos inputs, outputs y un programa especificable entre ambos, parece que sistemas no cognitivos de toda clase habrían de ser tenidos por cognitivos (Ibíd.: 191) y el reino de lo mental acabaría tornándose omnímodo (Ibíd.: 192). Es desarrollando este punto que hace una apreciación interesante desde el punto de vista de su posterior postulación del principio de conexión: la información computacional está sólo "en los ojos de los programadores y los intérpretes externos". Searle venía ya inclinándose hacia una concepción según la cual sólo dada la conciencia fenoménica, sólo partiendo de una mente como ésta que experimentamos, puede comenzar a hablarse con sentido de información, representación, intencionalidad y de mente en general, y en este sentido hablará en este artículo de "mentalidad genuina" y afirmará, sin ofrecer mayor explicación de lo que con ello quiere decir, que la distinción entre lo mental y lo no-mental "debe ser intrínseca a los sistemas" (Ibíd.: 191τ).

Otro punto destacable de esta segunda parte es que Searle, en su contrarréplica a la réplica del simulador de cerebros (Ibíd. 193-194) aplica su argumento al incipiente programa conexionista en ciencias cognitivas –téngase presente que faltaban seis años para que la Biblia del PDP (Rumelhart, McClelland et al., 1986; McClelland, Rumelhart et al., 1986) fuera publicada–.[62] Según esta réplica, podríamos obtener comprensión no

[62] Más recientemente, en Faigenbaum (2003), Searle se ha pronunciado nuevamente sobre el conexionismo en la misma línea crítica, apuntando que un enfoque realmente neurobiológico resultaría más provechoso para las ciencias cognitivas y, en paralelo a las últimas páginas del artículo de 1980 que venimos comentando, pronunciándose contra la metáfora del procesamiento de información e incidiendo en que "necesitamos distinguir entre el nivel de la información real, en el que nos hallamos ante procesos de

manipulando símbolos formales de forma serial, sino simulando computacionalmente, pero en paralelo, las operaciones del cerebro: "simulando las secuencias reales de descargas neuronales en las sinapsis".[63] Searle comienza por destacar que le parece una réplica alejada de lo que él entendía por IA fuerte, porque, en su concepto de la misma, una de sus características fundamentales consistía en la negación de que necesitemos saber cómo funciona el cerebro para entender el funcionamiento de la mente. Si para elaborar los programas de cómputo en los que Searle venía considerando que según la IA fuerte consistía la esencia de lo mental necesitamos saber cómo funciona el cerebro, el programa de la IA fuerte carece de sentido. El orden explicativo de la IA fuerte quedaría pues transfigurado por la lógica implícita en esta réplica, dado que mientras que la IA fuerte habría de plantear que el cerebro comprende porque instancia los programas adecuados, desde el punto de vista de esta réplica el programa sería el correcto en la medida en que fuera capaz de remedar adecuadamente determinados aspectos del funcionamiento neurobiológico (según Searle, como veremos, no los aspectos importantes por lo que a la capacidad de dicho funcionamiento de dar lugar a estados mentales toca).

En cualquier caso, esto es, a pesar de que la réplica del simulador de cerebros no se muestre –según lo antedicho– capaz de salvar a la IA fuerte de las consecuencias del argumento de la habitación china, Searle considera que dicha réplica no tiene, con todo, ningún objeto y que la situación que ella invita a imaginar, una supuesta descripción del programa conexionista en ciencias cognitivas, se ve afectada por las implicaciones del experimento mental de la habitación china exactamente igual que el programa simbolista clásico de IA fuerte. Así, propone Searle, la simulación en paralelo de la actividad cerebral sigue sin portar comprensión, por cercana que se halle al funcionamiento del cerebro, cosa que argumenta sugiriendo que en su habitación, su programa, su conjunto de reglas, podría consistir en una serie de instrucciones acerca de cómo manipular un complejo conjunto de pipas de agua conectadas con válvulas: al recibir los textos, configura el conjunto de pipas de forma que produce la secuencia de sinapsis apropiada para hacer que de la habitación salgan los textos chinos oportunos. Searle, hemos de entender, sigue superando el test chino de Turing, pero sigue igualmente sin comprender chino. El problema del simulador de cerebros (del programa conexionista en ciencias cognitivas), según Searle, es que simula el aspecto equivocado del cerebro: "la es-

pensamiento o algún otro estado mental intrínseco, y la forma *como-sí* o relativa al observador de hablar de "información", que es sólo una manera útil de hablar" (Faigenbaum, 2003: 62).

[63] Ciertamente, cabe poner en duda que esta frase conduzca a una comprensión cabal del sentido del programa conexionista en ciencias cognitivas.

tructura formal de la secuencia de descargas neuronales" en lugar de "las propiedades causales" del cerebro, "su capacidad para producir estados intencionales". En su planteamiento, "propiedades formales" y causales aparecen nítidamente diferenciadas (Searle, 1980: 194-195), y lo importante acerca de las operaciones del cerebro no es "la sombra formal proyectada por la secuencia de sinapsis, sino las propiedades reales de las secuencias" (Ibíd.: 198τ). Searle considera, pues, que la atención destinada unilateralmente a lo formal, a una *estructura abstracta* (Moural, 2003: 222) en lugar de a los poderes o propiedades causales del cerebro –unos poderes y propiedades, anotemos, de los que nada nos dice–, sentencia por igual al programa simbolista clásico y al conexionista en ciencias cognitivas.[64] La noción de lo "formal" vuelve a requerir aquí una reinterpretación –que Searle no ofrece– capaz de hacerla encajar en un nuevo contexto, más alejado esta vez de la noción tal y como fuera introducida al comienzo del artículo. Dicha noción, pues, es utilizada en este artículo en contextos diferentes, aludiendo en ocasiones a operaciones de tipo sintáctico[65] y en ocasiones a operaciones basadas en la *forma* (en el sentido de "aspecto" o "apariencia") de los símbolos, llegando Searle a proponer como intercambiables o equivalentes la hipótesis del sistema de símbolos físicos de Newell y Simons y la idea de que "la ejemplificación de un programa *formal* con los inputs y outputs adecuados es una condición suficiente, y de hecho constitutiva, de intencionalidad" (Searle, 1980: 195τ).[66] Este núcleo teórico es presentado además como la médula de la IA fuerte, de forma que lo que está en juego con la utilización de una noción de lo "formal" excesivamente liberal no es poca cosa –desde el punto de vista de los objetivos de la argumentación de Searle–. Sea como fuere, Searle considera funda-

[64] En su posterior respuesta a la crítica conexionista de los Churchland, que habían argumentado que el cerebro es de hecho un computador, pero muy diferente de los digitales, por cuanto opera en paralelo, no está programado de antemano y su principio rector no es la manipulación de símbolos formales pautada por reglas codificadas (Churchland & Churchland, 1990), Searle, en lugar de admitir que dicha propuesta se halla alejada de la IA fuerte tal y como él la comprende y ataca, entiende que la réplica de los Churchland no se aleja en exceso del marco de la IA fuerte tal y como él la interpreta y que el mismo experimento mental de la habitación china, modificado para que muchas habitaciones como la original funcionen en paralelo, bloquea las consecuencias que los Churchland pretendían extraer de su réplica (pues no hay motivos para suponer que en cualquiera de las habitaciones –neuronas– o en el conjunto de las mismas tenga lugar ninguna clase de comprensión). En su respuesta, Searle (1990d) falla al identificar la tradición simbolista con la conexionista y al obviar las diferencias entre ambas, cosa que más tarde corrigiera en Searle (1993b).

[65] La ambigüedad en este punto no es resuelta por alusiones posteriores en las que Searle dice cosas tales como que "la computación se define enteramente formal o sintácticamente" (Searle 2007b: 173τ).

[66] Las cursivas son nuestras. Aprovechamos para aclarar que Searle no se ha preocupado en escritos posteriores de especificar la noción de lo "formal" que utiliza en su argumento. Buen ejemplo de lo dicho lo ofrece un pasaje de Searle (1984b: 33) en el que sugiere, literalmente, que lo que significa que los programas computacionales se definan o especifiquen formalmente es que carecen de semántica. Añadamos que esta sentencia constituye, desde cierto ángulo, y no uno particularmente enrevesado, una estrepitosa *petitio principii*.

mental admitir que "las manipulaciones de símbolos formales [*formal symbol manipulation*] por sí mismas no tienen ninguna intencionalidad" (Ibíd.: 199τ): en ellas podemos encontrar sintaxis, dirá Searle, pero no semántica.

Esta última cita de Searle nos conduce al referido argumento teórico cuyas conclusiones serían idénticas a aquéllas a las que el experimento mental de la habitación china habría de conducir. Ciertamente, en el artículo de 1980 que venimos comentando dicho argumento se limita al contenido de la última cita (y a la sugerencia de las conclusiones del mismo en el resumen del artículo): Searle propone que las manipulaciones formales (ahora "formal" en sentido *sintáctico*, más allá pues de la interpretación ofrecida por Searle al principio del artículo) carecen de intencionalidad, y que con ellas no podemos ir más allá de la mera sintaxis. En escritos posteriores (Searle, 1984b: 39-41) ha desarrollado el argumento teórico aquí implícito del siguiente modo:

P1: Los cerebros causan las mentes. (Verdad de hecho).

P2: La sintaxis no es suficiente para la semántica. (Verdad conceptual).

P3: Los programas computacionales son enteramente definidos por su estructura formal o sintáctica. (Verdad por definición).

P4: Las mentes tienen contenidos mentales; específicamente, contenidos semánticos. (Hecho obvio).

C1: Ningún programa computacional es por sí mismo suficiente para proporcionar a un sistema una mente. (A partir de P2, P3 y P4).

C2: La actividad cerebral no puede causar la mente sólo en virtud de ejecutar un programa computacional (A partir de P1 y C1).

C3: Algo que pueda dar origen causal a la mente ha de poseer poderes causales al menos equivalentes a los del cerebro. (A partir de P1).

C4: Para cualquier artefacto dotado de estados mentales equivalentes a los humanos que podamos construir, la implementación de un programa computacional por sí mismo no será suficiente. Dicho artefacto habrá de tener, más bien, poderes causales equivalentes a los del cerebro humano. (A partir de C1 y C3).[67]

[67] Searle (1984c: 231-232; 1990d: 21-23) incluyen, en esencia, idénticas premisas y conclusiones. La versión más escueta –y asimismo la más explícita en lo tocante a la modalidad de su argumentación– puede hallarse en Searle (1989b: 701-702), en la que propone: no es el caso que los programas computacionales *necesariamente* impliquen mente, porque es *posible* tener programa sin mente (cosa esta última que el experimento mental de la habitación china debiera probar).

Una versión de su argumento más resoluta puede hallarse en Searle (1997a: 11-12 del original, 25 de la traducción):

1._Los programas son enteramente sintácticos.
2._Las mentes tienen semántica.
3._La sintaxis ni es lo mismo que la semántica ni es suficiente para ella.
Por tanto, los programas no son mentes.

A pesar de que este argumento se hallara sólo implícito en su artículo de 1980, Searle sugiere en la primera página del mismo que eso que aparecería en escritos posteriores como conclusión principal del argumento (es decir, el contenido de C1), es el resultado filosófico crucial alcanzado con el experimento mental de la habitación china, y, posteriormente (Searle, 1984c: 231), que el argumento es un resumen de lo que el experimento mental trata de mostrar. Sin embargo, más recientemente (Searle, 1990d: 21; 1994b: 546), Searle ha tendido a proponer que la función que cumpliría el experimento mental sería la de *apoyar* o *ilustrar* (son las nociones que, alternativa y respectivamente usa Searle en este punto, en Searle, 1990d y Searle, 1994b) P2, la cual, en la medida en que es tratada por el de Denver como una verdad conceptual (Searle, 1984b: 39, 1990d: 21) no se hallaría necesitada del sustento adicional que, desde esta interpretación, proporcionaría el debatidísimo experimento mental.[68]

Volviendo al artículo de 1980, Searle plantea al final del mismo que la característica de la IA que, precisamente, parecía hacer de ella una disciplina interesante desde el punto de vista del estudio de lo mental, la de la distinción entre programa y realización, resulta fatal para la disciplina. Parecía que con esa distinción podíamos estudiar lo mental en sí mismo sin recaer en la psicología introspectiva y avanzando hacia la autonomía de la psicología desde el respaldo que tal distinción ofrece al abandono del reduccionismo neurofisiológico. Sin embargo, propone Searle, esta ecuación según la cual la mente es al cerebro como el programa al hardware, trae consigo problemas mayores que los que venía a solucionar, entre los que destaca una excesiva atención a lo formal (dado que los programas, dice ahora, son *puramente formales*) a expensas del *contenido* (dice

[68] Searle, por otra parte, no se ha preocupado en las formulaciones posteriores de su experimento mental y su argumento de esclarecer las relaciones lógicas que habría entre ambos, limitándose, sencillamente, a presentar el uno detrás del otro.

en este sentido que lo intencional puede contener elementos formales pero que no se define por ellos, y pone el ejemplo de una creencia que es una y la misma con independencia de las formas sintácticas que la expresen), y una extraña recaída en un dualismo implícito que asume que la mente es independiente del cerebro y que el estudio del cerebro poco puede ofrecer a la comprensión de la mente.

Searle (1980: 187) plantea una cuestión crucial en este artículo: ¿en qué consiste eso en virtud de lo cual *comprendemos* y por qué no podríamos dárselo a una máquina, sea lo que sea? Bien, puede que Searle no ofrezca respuesta a la primera parte de esta pregunta (y, ciertamente, sería esperar demasiado de un breve artículo: todo lo que dice al respecto es que cierto tipo de organismo posee cierto tipo de estructuras biológicas que en ciertas circunstancias es causalmente[69] capaz de producir cierto tipo de estados mentales), pero su apuesta por lo que a la segunda toca no deja de ser fuerte y sus implicaciones de considerable magnitud: sea lo que quiera que sea eso en virtud de lo cual comprendemos, no podremos proporcionárselo a una máquina simplemente a base de procesos computacionales sobre elementos definidos formalmente: "ningún modelo puramente formal será nunca suficiente por sí mismo para la intencionalidad, dado que las propiedades formales no son por sí mismas constitutivas de la intencionalidad y dado que carecen por sí mismas de poder causal excepto, cuando son instanciadas [*ejemplificadas concretamente*, según la traducción española], el de producir la siguiente etapa del formalismo" (Ibíd.: 198τ).

> La capacidad causal del cerebro para producir intencionalidad no puede consistir en su ejemplificación concreta de un programa de cómputo, ya que para el programa que se prefiera es posible que algo lo ejemplifique y siga sin tener estados mentales. (Ibíd.: 204τ).

En definitiva, el experimento mental de la habitación china y el concomitante argumento teórico pretenden probar que el programa de investigación de la IA fuerte se halla, por principio, condenado al fracaso, dado que resultaría crucial para que dicho programa cosechara alguna clase de éxito que lo esencial acerca de lo mental pudiera ser descrito en términos de programas realizando alguna clase de procesamiento de información. Searle parece convencido de que la habitación china prohíbe esa posibilidad,

[69] Searle defiende aquí una postura en la que insistirá en escritos posteriores –especialmente en Searle (1992)– y que resulta un tanto indefinida: sólo poseyendo los poderes causales que posee el cerebro para producir conciencia e intencionalidad puede poseerse conciencia e intencionalidad.

prohibición ejemplificada en la imposibilidad de que algo cuente como mental sólo en virtud de constituir una instanciación del tipo *indicado* de programa computacional.

Searle, en su ataque al computacionalismo en la filosofía de la mente y ciencias cognitivas, ha sumado a lo apuntado en este apartado un nuevo argumento (Searle, 1990c; 1990e; 1992; 1994b; 1997a; 2007b) que cabe resumir en muy pocas palabras: la computación es un hecho relativo al observador (y por tanto no ontológicamente independiente de la existencia de éste), dado que los símbolos y la sintaxis lo son (porque no hay en el utillaje físico de las entidades que instancian propiedades sintácticas nada que las haga tales instancias más allá de la intencionalidad de sus intérpretes y usuarios: la sintaxis en sí misma no puede definirse en términos físicos, pues no es intrínseca a lo físico), y lo definitorio de la computación en tanto computación es exclusivamente sintáctico. Este punto de vista puede resumirse más aún indicando, con Searle (1990c: 636),[70] que no hay símbolos en la naturaleza, y que algo es un símbolo sólo cuando lo tratamos, consideramos o usamos como un símbolo (Searle, 2007b: 173). Partiendo de este nuevo argumento cabe asimismo negar que los procesos cerebrales sean intrínsicamente computacionales, dado que, de hecho, no existe nada intrínsicamente computacional (Searle, 1992: 225 del original, 230 de la traducción, no particularmente fiel al original en este punto). En este sentido, la crítica de Searle se extiende con el nuevo argumento de la IA fuerte a la neurociencia computacional e incluso a la neurociencia cognitiva al completo, si hemos de entenderla en el sentido de Churchland y Sejnowski (vid., para una resoluta panorámica, Churchland & Sejnowski, 1988) o en el implícito en las sucesivas ediciones de manuales y compilaciones de textos de neurociencia cognitiva que Gazzanigga viene dirigiendo desde mediados de los ochenta. ¿Por qué? Sencillamente porque el nuevo argumento incluye al anterior y va más allá: la IA fuerte sigue resultando refutada en tanto lo mental sigue sin ser cuestión de la ejecución de los programas computacionales correctos (en este punto el argumento sería ahora algo diferente: si no hay computación en la naturaleza, y si la mente es causada por el cerebro, en tanto que el cerebro no puede definirse como intrísecamente computacional en ningún sentido, la mente no puede ser tampoco en ningún sentido entendida como resultado de la ejecución los programas computacionales adecuados), pero el nuevo argumento se

[70] En este escrito Searle (1990c: 637) presenta sus conclusiones al respecto con una cautela –basada en sus dudas acerca de la posibilidad de una definición de computación que eluda su definición en términos literales de manipulación de símbolos– que desaparecerá en escritos posteriores.

opone también a una posición menos abstracta y más moderada, como la de la neurociencia cognitiva (a la que Searle no alude explícitamente) tal y como es habitualmente entendida (como el proyecto de descubrir y describir los algoritmos que operan en la actividad cerebral, así como la interrelación entre los mismos), es decir, el nuevo argumento minaría los cimientos de un programa de investigación basado en la idea de que al menos algunas de las operaciones del cerebro, para ser comprendidas por lo que a su relación con lo mental toca, han de ser conceptualizadas en términos computacionales. Más generalmente, el nuevo argumento se presentaría como un obstáculo a priori insalvable para un programa de investigación basado en la idea de que, para comprender al menos algunas rutinas neurofisiológicas, debe haber un nivel de descripción de esa actividad fisiológica en el que la explicación de la misma pueda alcanzarse detallando el tipo de procesos computacionales que de algún modo realiza, y aquí ya no hablamos de la imposibilidad de toparnos con fenómenos mentales tales como la comprensión sólo a base de manipulaciones de símbolos formales no interpretados, sino de la inviabilidad del proyecto de describir y explicar adecuadamente en términos neurocomputacionales procesos tales como, por utilizar un ejemplo del propio Searle (2007b: 172), el reflejo vestíbulo-ocular –vid., v. g., la amplia discusión de este reflejo desde un punto de vista neurocomputacional en la obra de referencia *The Computational Brain* (Churchland & Sejnowski, 1992: 353-378)–. La apuesta de Searle es ahora más ambiciosa. ¿Por qué? Sencillamente porque el argumento de la habitación china sólo lograría, si acaso lo hiciera, mostrar que la IA fuerte no puede lidiar con un determinado aspecto de alto nivel de la cognición: la comprensión lingüística. Sin embargo, aun cuando parece intuitivamente indudable que gran cantidad de criaturas no lingüísticas poseen capacidades cognitivas, una versión canina de la habitación china carecería de objeto. Pero la reformulación del anticomputacionalismo searleano basada en la idea de que la computación es algo dependiente del observador no pretende tirar por tierra las posibilidades explicativas del computacionalismo respecto de la capacidad cognitiva de la comprensión lingüística, sino sus posibilidades explicativas respecto de la cognición y lo mental en general. Searle se situa en una perspectiva esencialista: juega a la guerra de los paradigmas tratando de desterrar absolutamente todo lo que uno de ellos pudiera ofrecer en lugar de intentar hacer espacio para que algunas de las herramientas teóricas y experimentales gestadas en el seno del mimo puedan alcanzar a contribuir a nuestra comprensión científica de la mente. Del mismo modo que si algo mostraba el argumento de la habitación china ello no pasaba de ser que las siluetas (las *formas*) de los ideogramas no son sufi-

cientes para explicar la comprensión lingüística, la más reciente y ambiciosa apuesta de Searle vendría a mostrar que sólo los poderes causales (una noción que no explicita) pueden ocupar algún papel en nuestras explicaciones. Parece una conclusión un tanto inocua, pero no lo es. Todo lo que no sea "intrínseco a la física" queda desterrado. Pero, ¿podemos decir que sean "intrínsecas" a la física las funciones biológicas a que necesitan recurrir las explicaciones en biología? Bien cabe que el computacionalismo no sea *la* verdad acerca de la mente, como cabe igualmente que pueda contribuir a aumentar la profundidad y penetración de nuestro inquirirla mediante la integración de algunas de sus herramientas en alguno de los contextos experimentales o niveles explicativos a los que en el curso del mismo hayamos de recurrir. La argumentación de Searle, al margen de lo que quepa entender que hace al primer respecto, resulta enteramente estéril por cuanto no hace nada al segundo.

## 5._Crítica del planteamiento de Searle del problema de la conciencia

Aunque entendemos que el problema de la conciencia es un problema explicativo, habremos de centrar nuestra crítica del marco teórico de Searle en cuestiones ontológicas, y ello porque el mismo, a pesar de ser el más citado entre los elaborados dentro de la filosofía anglosajona de la mente, no ofrece nada en aquel sentido.

Los dos aspectos del planteamiento searleano con los que empezáramos nuestra exposición en este capítulo, el realismo ingenuo y la distinción entre epistemología y ontología, no son independientes: si podemos trazar, con Searle, una línea bien definida entre ontología y epistemología, esto se debe a que los hechos (la ontología) son previos e independientes de nuestras teorías (epistemología), y si esto es así, ello se debe a su vez a que existen una serie de hechos definitivos, incuestionables, evidentes e inapelables (realismo). Searle aplica este esquema realista a la mente y nuestras teorías acerca de la misma resultando así su enfoque caracterizable como un mentalismo ingenuo, según la etiqueta que él mismo utiliza (Searle, 1984b: 27; 1992: 28 del original, 42 de la traducción). De dicha aplicación dan cuenta numerosos pasajes de la obra de Searle (vid., v. g., el apéndice al segundo capítulo de Searle, 1992: 58 y ss. del original, 71 y ss. de la traducción). Por lo que toca a los estados mentales conscientes, en ellos, según Searle, no sólo se da ese carácter de anterioridad a nuestras teorías, sino también la imposibilidad de error respecto de los mismos, dado que, entiende, por lo que a ellos respecta no hay diferencia entre cómo son las cosas y cómo nos parecen, motivo por el

cual insiste (vid., v. g., Searle, 1989a: 203; 1992: 122 del original, 131 de la traducción; 1997a: 112 del original, 118 de la traducción) en que en el caso de la conciencia la única realidad es la apariencia. "En consecuencia, nuestro autor cree que con respecto a la mente hay una serie de hechos evidentes e incuestionables, de los que cualquier teoría de la misma debe partir y, en ningún caso, puede negar" (Guerrero del Amo, 2001: 304). En este marco, Searle pretende sostener a la vez el señalado realismo ingenuo y el relativismo conceptual que ha defendido explícitamente en diversos escritos (vid., v. g., Searle, 1991a; 1995a): la realidad de lo mental es una y la misma y así la captamos, pero luego la describimos, conceptualizamos o enunciamos en diferentes términos en función del marco conceptual en el que nos situemos. La propuesta de Guerrero del Amo en este punto es que la distinción, tajante y clara, entre captación y enunciación en el ámbito de lo mental de la que Searle parte resulta inviable y no trae consigo, en cualquier caso, las consecuencias que Searle busca en la misma, porque esos hechos indudables a los que hemos de atender para formular y contrastar teorías, en el propio acto de enunciarlos, pasarían a contar como verdaderos, exclusivamente, dentro del marco conceptual en que hayan sido enunciados, y no con la absoluta claridad y distinción que Searle supone. En su marco teórico, pues, nos encontramos con hechos mentales absolutamente evidentes y que todos deberíamos reconocer, pero que no pueden ser públicamente establecidos si no es de un modo relativo (a determinado marco conceptual), con lo cual careceríamos de un criterio intersubjetivo para la valoración de teorías acerca de la mente y los fenómenos mentales (Guerrero del Amo, 2001: 305). La clave de esta crítica estriba así en la imposibilidad de separar aquellos hechos mentales indubitables de los marcos conceptuales en que son enunciados y en la resultante imposibilidad de tratar públicamente dichos hechos indubitables fuera de esos marcos conceptuales de cara a la contrastación de hipótesis y teorías. En el caso de Searle, el marco conceptual en que se mueve es el del sentido común, de modo que los hechos indudables a los que apela sólo contarían como hechos inconcusos acerca de lo mental dentro de este marco, de forma que para aceptarlos como tales hemos de aceptar también dicho marco conceptual. A Searle, además, no sólo le parece evidente la verdad de nuestras intuiciones de sentido común acerca de lo mental (mentalismo ingenuo), sino que el conjunto de su planteamiento filosófico sobre la mente y la conciencia le merece la misma opinión: "Searle takes the truth of what he has to say about the mind as being absolutely obvious and uncontroversial to any adequately educated member of contemporary western intelligentsia" (Corcoran, 2001: 310).

No profundizaremos en críticas de carácter epistemológico como la emprendida por Guerrero de Amo (a pesar de que *abundemos* –acepción cuarta de esta entrada en la vigésima segunda edición del diccionario de la lengua española de la RAE– en ella), sino que nos apoyaremos en el carácter ingenuo de las asunciones searlenas en que la misma se basa para criticar la ontología de la mente defendida por Searle, es decir, su naturalismo biológico. Partir del referido marco conceptual del sentido común da lugar no sólo a problemas epistemológicos como el denunciado por Guerrero del Amo, sino asimismo a problemas de corte ontológico derivados de dicho marco conceptual y que tendrían, en último término, origen en la pretensión de Searle de hacer compatibles sentido común y cientismo,[71] y, concretamente, en su intención de salvar a la vez un mentalismo ingenuo antirreduccionista, según el cual lo mental es diverso de su substrato neurofisiológico (dado que si sólo hablamos de neurofisiología perdemos la ontología de primera persona en la que insiste Searle), y una suerte de ontología fisicalista según la cual lo mental no es nada por encima o más allá de la neurofisiología (nada más allá de un rasgo físico del cerebro), de forma que, en conformidad con la misma, y sin ir más lejos, no cabría seguir hablando, como en la frase anterior, en términos de "*substratos* neurofisiológicos". Hacer justicia a ambos conjuntos de requerimientos lleva a Searle a negar en la misma página (Searle, 2002b: 60) que la conciencia sea idéntica a su base neurofisiológica ("it does not follow that consciousness is nothing but neuron firings")[72] y que la conciencia sea algo distinto de su base neurobiológica ("I deny that (...) consciousness is something distinct from its neurobiological base").

En opinión de Searle, por otra parte, quien considere que su intento de maridar ontología fisicalista –Searle insiste en que los fenómenos mentales son fenómenos biológicos totalmente normales, y éstos, a su vez, fenómenos físicos completamente ordinarios– con mentalismo ingenuo antirreduccionista resulta de algún modo problemático es

[71] Searle viene considerando hace décadas (vid. Searle, 1984b: 13) este problema de la conciliación de nuestra concepción científica contemporánea del mundo y nuestra autocomprensión de sentido común. Así, recientemente ha dedicado un artículo (Searle, 2006) a comentar diferentes aspectos de este problema que, con Sellars, podemos denominar *choque de imágenes* (Sellars, 1962: 25) –la conciencia, la intencionalidad, el lenguaje, la racionalidad, la libertad, las instituciones sociales, la política y la ética son los ocho aspectos acerca de los cuales Searle ofrece en el artículo citado algunas apreciaciones relacionadas con la tensión entre cientismo ("basic facts", en palabras de Searle) y autocomprensión de sentido común ("certain conception we have of ourselves", en los términos que Searle emplea en el artículo)–. La conclusión alcanzada por Searle en este artículo que más nos interesa destacar aquí es la siguiente: la posibilidad de *naturalizar* cada uno de los aspectos que discute brevemente en su artículo no implica que ellos hayan de ser reductibles o eliminables, ni que carezcan de un verdadero carácter intrínseco (Searle, 2006: 105) –venga esto a significar lo que venga a significar, Searle no lo aclara.

[72] Comparar con la afirmación (en la página siguiente del artículo citado): "consciousness is nothing but a neurobiological process" (Searle, 2002b: 61).

presa del dualismo conceptual tradicional que presenta las nociones de lo físico y lo mental como contrapuestas y como poseyendo extensiones mutuamente excluyentes. Como veremos, no es sencillo valorar la medida en que Searle evita con éxito el tipo de dualismo conceptual que denuncia, como tampoco es sencillo valorar la medida en que con semejante acusación lograría evitar las dificultades que origina su planteamiento. De cara a comenzar a exponer el sentido en que valorar esas medidas es, en efecto, tarea complicada, será útil volver brevemente sobre el artículo de Guerrero del Amo con el que iniciábamos esta sección crítica. En el mismo se defiende brevemente que Searle arrastra un dualismo implícito (Guerrero del Amo, 2001: 311 y ss.) que surge de su incapacidad para hacer compatibles sentido común y cientismo, dado que, en último término, no lograría sostener a la vez la concepción de sentido común de lo mental que defiende (y que es, según Guerrero del Amo, la que comporta un dualismo implícito del que Searle no logra desprenderse a pesar de su insistencia en presentar lo mental como no contrapuesto a lo físico) y la concepción científica contemporánea del mundo de la que –a pesar de sugerir que matizada– no desea distanciarse. Por otra parte, Guerrero del Amo propone en el artículo citado, sumándose a Stich (1987), Nagel (1993a), Kim (1995) o, posteriormente, y en el ámbito de la filosofía de la mente española, a José Hierro-Pescador (2006: 79), que Searle no logra convencer de que su postura no sea la de un dualista de propiedades. Por nuestra parte, entendemos que Searle, efectivamente, no logra desprenderse de cierto tipo de dualismo implícito, pero no entraremos a discutir si cabe salvar su propuesta de la frecuente acusación de dualismo de propiedades. El medio más eficaz a tal fin consistiría en plantear que si entendemos el dualismo de propiedades como una concepción de lo mental según la cual existen algunas propiedades, las mentales, que son distintas de las propiedades físicas, y si tenemos en cuenta que lo que Searle sostiene es que todas las propiedades mentales son, en determinado nivel de descripción, físicas, y que algunas propiedades físicas son, en determinado nivel de descripción, mentales, entonces, la señalada posibilidad de salvar la ontología de Searle de dicha acusación habitual no parece inalcanzable (vid. Corcoran, 2001: 311-312). Puede, pues, que el naturalismo biológico searleano difiera del dualismo de propiedades por cuanto quiere seguir interpretando las propiedades mentales como propiedades físicas, a pesar de presentar a las primeras como radicalmente diferentes de –y de hecho ontológicamente irreductibles a– las primeras. Puede. No entraremos a discutirlo. Sólo apuntaremos que Searle (2002b: 60) entiende que su aproximación causal a la relación entre conciencia y neurobiología marca la distinción decisiva entre su postura y la propia del

dualismo de propiedades, dado que considera que, al presentar a la conciencia como causalmente reductible, se aleja de cualquier forma de dualismo en el seno del cual la conciencia aparezca como un extra de la neurobiología diverso de ésta. No obstante, precisamente acerca de dicha aproximación causal y su idoneidad nos veremos obligados a hacer algunas observaciones no particularmente halagüeñas. Sí entraremos, en cambio, a comentar el primer aspecto señalado: el dualismo implícito del que Searle no logra desprenderse. Su marco teórico sigue arrastrando esta suerte de dualismo –de corte cartesiano en su estructura conceptual, aunque *no explícitamente* en su ontología (vid. Guillot, 2010: 108 y ss., especialmente 115-121; Shani, 2007; 2008: 301-303 para dos diferentes cartesianizaciones del planteamiento searleano)–, principalmente porque Searle no logra concretar el sentido en que su naturalismo biológico se encuentra en una situación favorable de cara a superar el dualismo conceptual que pretende evitar. "Físico", dice Searle, no ha de entenderse como contrapuesto a "mental": sus extensiones no han de ser necesariamente diversas y excluyentes, pues "el hecho de que una característica sea mental no implica que no sea física, y el hecho de que sea física no implica que sea no mental" (Searle, 1992: 14-15 del original, 29 de la traducción). Con todo, Searle recurrirá finalmente a divisiones excluyentes y equivalentes en lo esencial a las que intenta eludir, y seguirá necesitando caracterizar a los estados mentales en términos de ontología subjetiva o existencia para la primera persona y contraponiéndolos a tal efecto a hechos o fenómenos ontológicamente objetivos, pues sólo aquéllos –y en ningún caso éstos– poseen, como repite en numerosas ocasiones, una ontología de primera persona.

Si no debe entenderse a Searle como un teórico de la identidad (pues quiere salvar a los fenómenos mentales de su neta identificación con o su reducción ontológica a estados físicos) ni como un dualista de propiedades (pues pretende distinguir niveles de descripción para permitir que una y la misma *cosa* porte al tiempo propiedades físicas y propiedades mentales sin que éstas puedan concebirse como diversas), ¿qué tipo de postura es pues la que defiende Searle al sostener, a la vez, que los fenómenos mentales conscientes son hechos reales e irreductibles, y que estos mismos fenómenos son hechos biológicos enteramente ordinarios que tienen lugar en el cerebro como rasgos suyos? Una atenta lectura de la literatura searleana obliga a responder que dicha postura no pasa de constituir un curioso intento de hibridar dualismo y materialismo (Olafson, 1994: 255) rescatando lo que de intuitivamente cierto parece haber en uno y otro y derivando así en un dualismo biológico según el cual dentro de la clase de los fenómenos biológicos hay una subclase enteramente ordinaria de los mismos que denominamos

fenómenos mentales y que pueden equipararse a fenómenos biológicos tan corrientes y tan poco misteriosos como la digestión, la fotosíntesis o la mitosis (Searle, 1992: 1 del original, 15 de la traducción; 90 del original, 102 de la traducción), pero que, a pesar de este carácter no misterioso y ordinario, y a diferencia del resto de los fenómenos biológicos ordinarios, poseen una ontología de primera persona que trae aparejada una forma de relación epistémica con tales fenómenos enteramente diferente de nuestra relación epistémica con el resto de los fenómenos biológicos. Dualismo biológico, pues, dado que a pesar de la insistencia de Searle en el carácter ordinario de la clase de fenómenos biológicos que constituirían los mentales, los mismos se diferenciarían en su planteamiento del resto de los fenómenos biológicos tanto ontológica como epistemológicamente pues, como David Pineda ha señalado acertadamente, la conciencia es concebida por Searle, en tanto que subjetiva, no sólo como algo a lo que tenemos un tipo de acceso epistémico del que no disponemos en el caso de la mitosis o la digestión, sino como algo con una ontología totalmente diferente de la propia del resto de los fenómenos biológicos (Pineda, 1999: 155-156).

La subjetividad, como característica definitoria de la conciencia, como su modo de existir, es definida por Searle en términos de existencia para la primera persona (Searle, 1992: 94 del original, 106 de la traducción), es decir, como una característica de la conciencia que la convierte en privada. Diferentes observadores se hallan por tanto en distintas clases de relación epistémica con unos y los mismos fenómenos, pues cada organismo consciente se halla en una relación con sus estados conscientes (existen *para* él) que únicamente él mismo puede mantener: los estados mentales conscientes "no son igualmente accesibles para cualquier observador" (Ibíd.). Desarrollando esta perspectiva, Searle, a pesar de negarlo explícitamente (implícitamente ya en Searle, 1981b: 422; explícitamente en Searle, 1992: 98 del original, 110 de la traducción; 2007c: 97 del original, 121 de la traducción), y a pesar de rechazar el supuesto modelo espacial o visual que asocia con la doctrina del acceso privilegiado, se ve envuelto en sus redes, pues los estados mentales que, vía subjetividad, aparecen en su propuesta como fenómenos enteramente privados, quedan fuera del alcance del punto de vista de la tercera persona, y no sencillamente a su alcance pero desde un punto de vista diverso del de la primera persona. Ciertamente, nos hallamos aquí con una perspectiva difícil de conciliar con la *Weltanschauung* científica que Searle no pretende abandonar, pues sólo las peculiaridades ontológicas de la conciencia conducen a semejante tipo de relación epistémica con los objetos que sean el caso, y Searle no ofrece pistas acerca del modo de reformular dicha

*Weltanschauung* a fin de acomodar estas peculiaridades, ni acerca de la manera de ampliar el modelo científico de objetividad para dar cabida de forma epistemológicamente objetiva a objetos de descripción y explicación ontológicamente subjetivos, potenciales objetos de descripción y explicación científica únicos en su carácter, dado que en el resto de las áreas de la biología –disciplina a la que Searle alude como la crucial dentro de la señalada *Weltanschauung* por lo que a la conciencia toca– ningún objeto de descripción y explicación presentaría el señalado rasgo de la subjetividad ontológica, ni permanecería, por tanto, fuera del alcance de una inspección idéntica o análoga desde diferentes perspectivas, y menos radicalmente fuera del alcance de la inspección en tercera persona. La tensión entre dos conjuntos de requerimientos, provenientes del marco conceptual del sentido común y del mentalismo ingenuo antirreduccionista, por una parte, y de la pretensión de permanecer dentro de los márgenes de nuestra visión científica del mundo y asimismo dentro de los límites de la forma de fisicalismo a la que conduce el naturalismo biológico, por la otra, produce dificultades que no parecen resolubles desde dentro de la ontología de la mente que Searle destina, precisamente, a resolver semejante clase de dificultades. Como lo hasta aquí apuntado sugiere, la principal entre las dificultades y perplejidades a las que conduce la propuesta searleana consiste en que los estados conscientes, concebidos por Searle como rasgos físicos del cerebro, serían, a la vez, ontológicamente objetivos y ontológicamente subjetivos. Los mismos rasgos del cerebro serían al tiempo enteramente públicos, como consecuencia de la primera caracterización ontológica, y enteramente privados, como consecuencia de la segunda: nos hallaríamos así ante propiedades que son y no son objetivas en sentido ontológico (en función, según Searle, del nivel descriptivo que escojamos), y que, de este modo, caen y no caen fuera de las posibilidades de la observación en tercera persona. Siendo la conciencia, tal y como Searle la presenta, un fenómeno biológico enteramente ordinario, resulta extraordinario el modo en que viene a partir en dos el campo de los objetos de estudio de la biología pues, repitámoslo, no parece haber ningún otro objeto de las ciencias biológicas que comparta estas particulares características y que requiera de ellas (de las ciencias biológicas) los refinamientos, ampliaciones y rehechuras que dadas las mismas (dadas las señaladas particulares características) parecen imponerse a sus modelos descriptivos y explicativos. Surge así la pregunta acerca del modo de reformular dichos modelos descriptivos y explicativos para que den solvente cuenta del

hecho de que algunas propiedades de determinados sistemas biológicos no sean accesibles sino desde la primera persona[73] y sigan siendo, sin embargo, fenómenos biológicos enteramente normales que cabe de hecho describir como rasgos físicos –y por tanto, en tanto tales y en principio, es decir, hasta que no alcance Searle a explicarnos cómo introducir en nuestra noción de lo físico su noción de subjetividad ontológica (¿quizá recurriendo al panpsiquismo?), *enteramente* accesibles a la inspección en tercera persona–. Surge también la pregunta acerca del carácter distintivo de tales rasgos frente a otros rasgos no conscientes del sistema nervioso central: ¿qué explicación cabe ofrecer de dicha segmentación de lo biológico?; ¿cómo abordar, partiendo de los modelos explicativos de las ciencias biológicas, el hecho de que determinadas rasgos físicos del cerebro pasibles de tránsito objetivo al ámbito de lo epistémico, sean a su vez ontológicamente subjetivos y resulte por ello imposible permanecer respecto de los mismos en idénticas relaciones epistémicas al pasar de la primera a la tercera persona?; ¿cómo es posible que unos y los mismos rasgos físicos del cerebro, pasibles de idéntico tratamiento epistémico desde diferentes perspectivas, sean a la vez ontológicamente subjetivos y se caractericen por un acceso exclusivo desde la primera persona, es decir, se ofrezcan distintiva y hasta privativamente a una primera persona?; ¿cómo interpretar el hecho de que determinados rasgos físicos del cerebro sean radicalmente diversos de otros, por cuanto, siendo ambos enteramente físicos, sólo unos se hallen dotados de una ontología de primera persona y se encuentren así en relaciones epistémicas radicalmente diversas al pasar de la primera a la tercera persona?; ¿por qué algunas propiedades físicas del cerebro pertenecen a la primera persona –sólo *existen para ella*– resultando así privadas aun cuando puedan ser descritas como rasgos enteramente físicos del cerebro (de modo que con ellas nos hallaríamos a la vez con dos formas de existencia, con dos ontologías en función del nivel descriptivo que elegido: ¿hacemos así depender la ontología de cierta suerte de tarea de decisión semántica?), mientras el resto de las propiedades físicas del cerebro conservan exclusivamente su objetividad ontológica y se hallan así disponibles para la inspección pública en tercera persona ofreciendo una posibilidad de tránsito de la objetividad ontológica a la objetividad epistemológica que no resulta fácil imaginar

[73] Pues la subjetividad ontológica de Searle, acendrada cuanto se quiera de concomitancias epistemológicas, no puede significar nada muy alejado de esta inaccesibilidad, con independencia de la aceptación o rechazo del supuesto modelo espacial que implicaría la noción de acceso (vid. Searle, 1992: 70 del original, 81-82 de la traducción, un fragmento en el que la ontología de primera persona de la conciencia aparece vinculada a formas radicalmente diferentes de acceso epistémico para la primera y la tercera persona al punto que, desde la perspectiva de ésta, tan siquiera cabría decidir si un individuo en determinadas circunstancias es o no consciente).

cómo proyectar sobre la ontología de primera persona, dado que ella se halla vinculada con relaciones epistémicas radicalmente diversas al pasar de la primera a la tercera persona?; ¿por qué sólo algunos de los rasgos físicos del cerebro se hallan al alcance de la conciencia mientras el resto de los rasgos igualmente físicos del cerebro preservan sólo ontologías objetivas?; ¿cómo hemos de entender la existencia de unos rasgos físicos inaccesibles sino para una primera persona?; ¿de qué modo hemos de interpretar la nueva noción de lo físico que Searle estaría proponiéndonos, con la cual habría de hacerse de algún modo abordable desde la metodología científica un fenómeno accesible exclusivamente desde la primera persona?; ¿conducen el resto de las propiedades emergentes que Searle compara con la conciencia a estas dificultades?

Searle no ofrece pista alguna de la que partir de cara a hacer frente a cuestiones que, como las anteriores, surgen dentro un aparato conceptual que ha venido a sustituir las divisiones y contraposiciones tradicionales entre lo físico y lo mental por divisiones y contraposiciones más sofisticadas pero no menos paradójicas entre fenómenos ontológicamente objetivos y fenómenos ontológicamente subjetivos, los cuales, además, conviven en su planteamiento de tal modo que habrían de ser atribuibles al mismo tiempo a unos y los mismos rasgos de determinada entidad en determinadas circunstancias, es decir, de tal modo que unos y los mimos rasgos poseerían, a la vez, ontologías objetivas o de tercera persona y subjetivas o de primera. Las categorías polarizadas de lo que Searle denomina "dualismo conceptual" serán en su propuesta sustituidas por un dualismo en los conceptos de corte no menos tradicional: las divisiones y contraposiciones entre lo subjetivo y lo objetivo y entre la primera y la tercera persona se mantienen y ocasionan idénticos problemas y perplejidades que las categorías que Searle pretende abandonar con su conato de superación del dualismo conceptual, el cual, desde este punto de vista, afectaría no sólo a los dualistas del pasado y los materialistas del presente, sino también a su naturalismo biológico. En este sentido, Searle, ha llegado a proponerse, "sencillamente ha sustituido un dualismo por otro" (Moreland, 2008: 55τ), y así, al valorar la medida en que logra superar el dualismo conceptual que denuncia como omnipresente, no resulta injusto afirmar que "el naturalismo biológico no ha conseguido zafarse del todo de la tradición a la que con tanta vehemencia critica" (Pérez Chico, 1999: 128), y no sólo esto, sino que, además, a pesar de intentar tratar a los estados mentales como fenómenos biológicos, "no importa lo mucho que Searle insista en negarlo, ha de ser alguna forma de dualismo" (Ibíd.: 129), dadas las implicaciones que trae consigo la postulación de un ámbito de la realidad, el de lo mental, diverso del resto de

ámbitos y, de hecho, ontológicamente irreductible. La insistencia en la existencia de áreas irreconciliables de nuestra ontología no resulta, en cualquier caso, coherente con la invitación de Searle a no concebir la realidad como conformada por dos esferas contrapuestas, las de lo mental y lo físico.

Otro problema ontológico que acarrearía el *dualismo biológico* searleano implicaría a la concepción de la conciencia como propiedad emergente. Searle propone que, al igual que sucede con propiedades sistémicas emergentes como la liquidez (macropropiedades), no podemos figurarnos que un sistema haya de portar o dar origen a estados mentales conscientes atendiendo meramente a la disposición o configuración espacial de sus partes componentes, sino que sería necesario atender a las interacciones causales entre dichas partes componentes. La conciencia es causalmente emergente, nos dice. Pero el problema es que este emergentismo causal en el que se apoya el naturalismo biológico de Searle no acaba de explicar el tipo de relaciones habidas entre macropropiedades y micropropiedades. Searle hace referencia a las relaciones causales entre el macronivel de la conciencia y el micronivel fisiológico en términos ciertamente vagos. Una vez ha propuesto que la conciencia emerge causalmente, ¿a qué tipo de relaciones causales se supone que debemos atender cuando se nos dice que no es suficiente con tener en cuenta la disposición espacial de los microelementos, sino que para comprender el tipo de emergencia protagonizado por la conciencia hay que prestar atención a relaciones causales? ¿Se trataría de relaciones causales entre los propios microelementos, o entre la conciencia entendida, como propone, como una macropropiedad y aquéllos? Parece que Searle quiere referirse al segundo tipo de relaciones causales… ¡cuando se da el caso de que habla explícitamente de las primeras! (Searle, 1992: 112 del original, 122 de la traducción). Pero lo verdaderamente grave es que no sabemos (y Searle no nos lo dice) cómo salir del ámbito de las micropropiedades y dar el salto a la conciencia entendida como macropropiedad al describir o explicitar el pertinente funcionamiento causal neurofisiológico: ¿podríamos atender a todas las relaciones causales, digamos, horizontales (entre microelementos), sin hacer mención de la conciencia? Usemos el ejemplo de la solidez, una de las propiedades emergentes que Searle presenta como causales, esto es, como equiparables al tipo de emergencia relevante por lo que a la conciencia toca.[74] Una relación exhaustiva de todas las relaciones causales entre los micro-

[74] No entendemos demasiado bien la propuesta de Searle en este punto, porque, al contrario de lo que sugiere, el rasgo definitorio de la clase de agregación de la materia que denominamos solidez es un rasgo estructural que tiene que ver con la disposición espacial de los átomos y las moléculas: podemos bajar al

elementos que compongan un objeto sólido no hará ninguna mención de la solidez, y de hecho presentará al objeto como algo prácticamente vacío: hablaremos de intercambios de electrones entre los átomos que integren las moléculas del objeto y fenómenos por el estilo. La solidez aparecerá en otro nivel: al describir, partiendo de la constitución de nuestros sistemas sensoriales y de nuestra forma de relacionarnos con el mundo (no demasiado que ver, pues, con el modo de existir de lo descripto o con la pertinencia ontológica de los referentes de los términos que usemos en nuestras descripciones), relaciones *macro-macro* (no micro-macro, como pretende Searle a pesar de que acabe sólo por hablar de relaciones micro-micro, y asumiendo que tenga sentido hablar de este modo) entre el objeto que sea el caso y otros objetos incapaces de atravesarlo (téngase presente que podría ofrecerse una explicación de esta situación sin salir del micronivel). Entre los elementos del micronivel hay una plétora de relaciones causales horizontales que pueden explicar todas las propiedades que sean el caso obviando nuestro discurso acerca de un macronivel y que, así, parecen no importar por lo que respecta a la propuesta de Searle, pues lo que él busca es un tipo de relación causal vertical mediante la cual salimos enteramente del micronivel. En este punto surgen inevitablemente las dudas acerca de la existencia y, en tal caso, acerca de la forma adecuada de tratar ese tipo de relaciones causales verticales, pues Searle se muestra más bien escueto en este punto y siquiera encuentra problemática la idea de que quepa la posibilidad de que la introducción de *todas* las relaciones causales pertinentes entre elementos del micronivel no tenga que sacarnos necesariamente del micronivel.

Si Searle propone, pues, que la conciencia emerge causalmente, hay que entender que existe un tipo vertical de causación que va de los elementos del micronivel fisiológico al macronivel mental. Searle asegura que no hay nada extraño o misterioso en semejante tipo vertical de causación (Ibíd.: 126 del original, 135 de la traducción). Pero

---

micronivel, fijarnos en la disposición espacial de los átomos y las moléculas, comprobar que se hallan dispuestos según una estructura tipo red cristalina y decidir que se trata de un sólido (no porque podamos ver o experimentar su solidez dentro del propio micronivel, sino porque sabemos por los libros de texto que ese tipo de disposición espacial de los elementos en el micronivel corresponde aproximadamente con el tipo de disposición espacial de los elementos del micronivel que tendrían los objetos que llamamos sólidos en el macoronivel). Searle (1992: 111 del original, 121 de la traducción), no obstante, dice que hay que atender, también aquí, también en el caso de los estados de agregación, a las interacciones causales más allá de la mera disposición espacial. El extremo en litigio en esta nota al pie no afecta a la crítica que desarrollaremos. Así, quede como nota al pie, pero añadamos un par de preguntas: ¿cuántos microniveles contendrá el micronivel?; ¿cabrá distinguir en ellos de forma tajante lo causal de lo espacial?; en último término, ¿qué motivos tenemos para concebir como causal la relación que Searle propone entre sus más bien poco definidos reinos de lo micro y lo macro cuando todo parece apuntar a una relación constitutiva? "Los microelementos de Searle en el ejemplo de la solidez no causan la solidez (…), son la solidez misma. (Mora Teruel, 2001: 142).

cuando nos invita a concebir la conciencia como una propiedad causalmente emergente como la solidez o la liquidez, y asegura que puede ser enteramente explicada atendiendo a las relaciones causales horizontales habidas entre los elementos del micronivel neurofisiológico, es decir, *dentro del propio micronivel* –tal y como evidencia la afirmación de Searle según la cual "the existence of consciousness can be explained by the causal interactions between elements of the brain at the micro level" (Ibíd.: 112 del original, 122 de la traducción)–, lo que realmente está pidiéndonos es que *imaginemos* un tipo vertical de causación del micro al macronivel, y que abandonemos mediante dicho tipo vertical de relación causal el micronivel camino del macronivel. Con todo, independientemente de la ambigüedad presente en los textos de Searle entre un supuesto tipo vertical de causación y uno horizontal, es decir, independientemente de la ambigüedad con la que Searle presenta la forma de causación vertical a la que recurre aludiendo de hecho al tipo habitual de causación horizontal intranivel, cabe preguntar cómo haremos dicho camino, es decir, cómo abandonaremos el micronivel en dirección al macronivel mediante el aludido tipo vertical de relación causal. Searle no dice nada al respecto, y podría buscar auspicio en la frontera interdisciplinaria alegando que no es él quién deba ofrecer respuesta a semejante clase de pregunta, dado que sólo a disciplinas experimentales competería hacerlo. Pero, ¿cómo sabremos cuándo hemos abandonado el micronivel en nuestro recuento de las interacciones causales habidas dentro de él?; ¿cómo sabremos cuándo, en nuestra consideración de las relaciones causales pertinentes, abandonamos el ámbito de las relaciones causales horizontales intranivel para adentrarnos en el de las relaciones causales verticales internivel que Searle necesita? Tampoco nos ofrece Searle los medios para abordar estos interrogantes, y de hecho se muestra ambiguo y escueto al hablar de las apuntadas relaciones causales: ni siquiera ofrece razones concluyentes en defensa de la tesis de acuerdo con la cual existiría de hecho un tipo de causación vertical micro-macro como el que requiere su propuesta, sino sólo ejemplos que pueden interpretarse de diferentes modos: véase cuánto nos ha costado darle la vuelta a su ejemplo de la solidez para mostrar que hablar de relaciones causales entre un supuesto micronivel y un supuesto macronivel resulta, cuando menos, incierto: ¿qué se supone que podrían causar en un macronivel los elementos del micronivel cuando sólo encontramos relaciones dentro del propio micronivel que ya explican todo lo que cabe explicar, tal y como el propio Searle admite al señalar que la solidez no tiene poderes causales adicionales a los presentes en su base molecular (Searle, 2002b: 61)?. En esta ocasión no encontrará asilo en ninguna jurisdicción, dado que se trata de problemas ín-

sitos en su propio marco conceptual que nos llevan de vuelta a embrollos conceptuales de tipo dualista. ¿Ha superado Searle el dualismo conceptual que critica o lo ha sustituido por uno nuevo y aquejado de idénticos problemas, sólo que esta vez en términos de dualidad micro/macro en lugar de en términos de dualidad físico/mental? Podemos resumir este aspecto de nuestra interpretación del naturalismo biológico searleano en términos de dualismo biológico señalando que la solución al problema mente-cerebro que Searle ha venido proponiendo como obvia desde 1983 (vid. Searle, 1983: cap. 10) no deja de acarrear problemas casi idénticos a los que venía a solucionar, pues apuntar que un tipo de causalidad abajo-arriba (bottom-up) funciona de tal modo que el comportamiento de elementos en el micronivel fisiológico da lugar a (o *causa*) fenómenos conscientes en un macronivel de propiedades sistémicas mentales resulta poco explicativo cuando no tenemos clara la forma de distinguir ese tipo de causalidad vertical del tipo de causalidad horizontal que podemos encontrar analizando las relaciones entre elementos en el micronivel; cuando Searle, de cara a orientarnos, habla de dicha causalidad vertical en términos de causalidad horizontal argumentando, como veíamos, que "la existencia de la conciencia puede ser explicada por las interacciones causales entre elementos del cerebro *en* el micronivel" (Searle, 1992: 112 del original, 122 de la traducción),[75] y cuando no podemos estar seguros de que hablar de macropropiedades explicadas por un tipo de causalidad vertical abajo-arriba resulte *ontológicamente* lícito y exceda el ámbito de la economía léxica. Para esclarecer este último punto sólo es necesaria una atenta consideración de lo que está ya implícito en nuestra forma de dar la vuelta al ejemplo searleano de la solidez como propiedad causalmente emergente. Para Searle resulta obvio que la conciencia existe de un modo enteramente distinto a ése en que existen los billetes de lotería: éstos sólo existen en cuanto tales dadas nuestras creencias y prácticas. La conciencia no: ella existe con independencia de creencias, gustos, actitudes, prácticas, etc. No obstante, cabe dudar que la liquidez, otro ejemplo de propiedad causalmente emergente que Searle equipara en su forma de emergencia a la conciencia, exista con independencia de prácticas o actitudes, es decir, que exista en el mismo sentido en que existe el agua o la nitroglicerina (también líquida a temperatura ambiente) y se encuentre en una situación ontológica, a diferencia de las cualidades secundarias, enteramente autónoma. Dadas nuestras capacidades perceptivas y nuestras prácticas científico-técnicas, distinguimos tres estados de agregación de la materia: los

[75] Las cursivas son nuestras.

estados de agregación varían de tal modo que aparecen casos límite en función de las condiciones de presión y temperatura que no casan demasiado bien con términos del vocabulario cotidiano como «líquido» o «gaseoso», y podríamos de hecho distinguir más de tres formas de agregación entre la red cristalina de los sólidos y la libertad molecular de los gases si lo necesitáramos o las percibiéramos como obvias.[76] Searle lleva tan lejos como nosotros su postura respecto de la formulación de categorizaciones ontológicas y propone (Searle, 2002b: 59) que se trata siempre de un acto relativo a intereses. No se entiende cómo podrá entonces tratar de defender la diferencia entre la ontología de la conciencia, ajena en su propuesta a la presencia de actitudes, gustos, preferencias, prácticas o intereses, y la ontología de los goles en un partido de fútbol. ¿Será pues, tan obvio como Searle (Ibíd.: 61) pretende que tanto las macropropiedades como los microelementos forman parte de "la estructura causal del universo" cuando no encontramos por ninguna parte (ni Searle nos la muestra) la línea divisoria que nos indique: "a partir de aquí, los efectos de las relaciones causales entre elementos del micronivel ya no son efectos dentro del propio micronivel, sino efectos en una nueva esfera ontológica tan legítima, genuina y tan causalmente autónoma, real y eficaz como la anterior: la de las macropropiedades"? ¿Es realmente una relación de causación la habida entre microelementos y macropropiedades, por ejemplo, entre la estructura molecular y la solidez de un objeto? En resumidas cuentas, y a pesar de que Searle (1992: 126 del original, 135 de la traducción) sugiera que no hay nada misterioso relacionado con el tipo vertical de causación del que venimos tratando, la naturaleza de la misma y la distinción entre ella y el tipo ordinario de causación (al que, a falta de un mejor apelativo, venimos denominando *horizontal*) resultan desconcertantes y generan problemas en el aparato conceptual elaborado por Searle que no parecen ir a disiparse con simplemente afirmar "esto no es en absoluto misterioso: ¡la causación bottom-up es ubicua en la naturaleza!".

La ontología de la mente de Searle arroja más problemas relacionados con su tratamiento de la causalidad. Acabamos de sugerir que Searle no ofrece argumentos sólidos en defensa de una tesis central desde el punto de vista de su ontología de la conciencia: la tesis según la cual existe, efectivamente, un tipo de causación vertical (bottom-up o abajo-arriba) responsable de la existencia de propiedades macro y claramente diferenciable del habitual tipo de causación horizontal. La distinción, en resumidas

[76] "It is useful for our brains to construct notions like solidity and impenetrability, because such notions help us to navigate our bodies through a world in which objects –which we call solid– cannot occupy the same space as each other" (Dawkins, 2006: 368 del original, 426 de la traducción).

cuentas, y contra lo que Searle pretende de ella, puede tener más que ver con nuestras formas de describir que con la forma de existir de las cosas que describimos. Entendemos, por otra parte, que tampoco ofrece Searle, ya no argumentos, sino una clara explicación de qué defiende cuando sostiene, a la vez, 1) la clausura causal del mundo físico (Searle, 2002b: 61; 2004a: 136), 2) la reductibilidad causal de la conciencia, y 3) la irreductibilidad ontológica de la conciencia. Una posible lectura de las implicaciones de la conjunción de estas tres tesis consistiría en afirmar que con ellas Searle pretende defender (por 1 y 2) la inexistencia de epifenómenos, y, a la vez, la conciencia como una excepción a tal regla, es decir, como un epifenómeno (por 3). A la misma conclusión llega David Pineda (aunque partiendo de las tesis de Searle según las cuales la conciencia superviene causalmente pero no es causalmente reductible): "Si la relación entre los estados mentales y los estados físicos es la que nos propone Searle entonces los primeros son epifenómenos de los segundos, en contra de lo que el propio Searle pretende" (Pineda, 1999: 159). Y a la misma conclusión llegamos si partimos, con Palmer (1998: 163-164), y con el propio Searle (2002b: 60), de la idea según la cual, dada una reducción causal, los poderes causales de los fenómenos reducidos no pasarían de ser, exhaustivamente, los poderes causales de los fenómenos reductores. No debería resultar una lectura extraña, a pesar de los esfuerzos de Searle para eludir el epifenomenalismo. Después de todo, la tesis 1 tiene un fuerte carácter ontológico: todo lo existente en el mundo físico puede, en último término, definirse enteramente atendiendo a criterios causales. Añadir a esta idea las contenidas en las tesis 2 y 3 nos conduciría a lo siguiente: a pesar de que la conciencia pueda definirse atendiendo a criterios causales (tesis 2), esa definición nos dejaría (tesis 3) sin lo que resulta esencial de la conciencia (Searle, 1992: 121 del original, 131 de la traducción); luego, lo esencial de la conciencia no puede capturarse en ninguna relación de hechos y procesos meramente causales. De aquí a cierta suerte de epifenomenalismo no parece haber más que un paso. Searle, por otra parte, se encuentra en una posición extraña al suscribir a la vez estas tres tesis, porque aceptar el principio del cierre causal del mundo físico y proponer al mismo tiempo que existen fenómenos biológicos tan comunes como cualquier otro, pero que ontológicamente se diferencian del resto de los fenómenos que tienen lugar en el mundo físico en que escapan necesariamente a cualquier aproximación meramente causal, parece conducir a una contradicción: los estados conscientes, según Searle, pueden, como todo el resto de estados de cosas, redefinirse completamente en términos causales, pero, a diferencia del resto de estados de cosas, exceden tal redefinición al caer, esencialmente,

fuera de la misma, dado que semejante redefinición, según Searle, dejaría fuera lo característico de los estados conscientes: su ontología de primera persona. Atendiendo al planteamiento de Searle, pues, los estados conscientes pueden y no pueden redefinirse en términos causales: un inventario de la historia causal completa del universo (un inventario que, por el principio del cierre causal del mundo físico, no dejaría fuera nada, contemplaría *todo*) contendría (tesis 2) y no contendría (tesis 3) a la conciencia. No explotaremos las vías críticas que abren las apreciaciones sumariamente realizadas en este párrafo. En lugar de ello, nos fijaremos brevemente en los problemas que trae para Searle el deseo de oír misa y repicar al mismo tiempo, es decir, el deseo de sostener a la vez que la conciencia es reductible e irreductible.

¿Qué tiene Searle en mente al argumentar que la conciencia es causalmente reductible pero ontológicamente irreductible? Al afirmar que la conciencia es causalmente reductible Searle no estaría diciendo nada más allá de lo siguiente: la existencia de la conciencia y la de sus poderes causales se explican *enteramente* por los poderes causales del cerebro. Al afirmar que la conciencia es ontológicamente irreductible estaría diciendo, por otra parte, que al realizar la más minuciosa entre las explicaciones de la conciencia realizadas en términos neurofisiológicos que quepa imaginar estaríamos dejando fuera lo esencial de la conciencia: su subjetividad, su ontología de primera persona. Estas dos afirmaciones, según Searle, no entrarían en conflicto, y además, la irreductibilidad ontológica de la conciencia no traería consigo implicaciones dualistas de ninguna clase.

> Causal reduction does not necessarily imply ontological reduction, though typically where we have a causal reduction as in the case of the liquidity, solidity and colour we have tended to make an ontological reduction. But the impossibility of an ontological reduction in the case of consciousness does not give it any mysterious metaphysical status. Consciousness does not exist in a separate realm and it does not have any causal powers in addition to those of its neuronal base any more than solidity has any extra causal powers in addition to its molecular base. (Searle, 2002b: 62).

No obstante, repasando la obra de Searle, encontramos un solo argumento en defensa de lo contenido en las primeras palabras de la última cita, es decir, en defensa de la tesis según la cual una reducción causal no implica necesariamente una reducción ontológica. ¿Por qué cree Searle, pues, que la conciencia es causalmente reductible aunque ontológicamente irreductible? Según Searle, dado el patrón de nuestras reducciones, una vez realizadas las reducciones causales pertinentes, procedíamos a realizar las co-

rrespondientes reducciones ontológicas. Esto, según Searle, es posible tanto en el caso de las propiedades secundarias como en el de las primarias, y en el caso de estas segundas, tanto en sus aspectos objetivos como subjetivos: una vez que comprendemos que todos los rasgos superficiales de un fenómeno, ya sean subjetivos (como la experiencia del calor) u objetivos (como la impenetrabilidad de un objeto sólido), son "efectos de la cosa real" (Searle, 1992: 120 del original, 129 de la traducción), redefinimos esos rasgos en términos de su base causal y realizamos una reducción ontológica con arreglo a la cual sostenemos algo que puede expresarse de un modo bien sucinto: aquellos rasgos superficiales no *eran* en verdad nada más que estos fenómenos causales subyacentes. Lo mismo podría hacerse con la conciencia si insistiéramos en ello, dice Searle (Ibíd.: 121 del original, 131 de la traducción), pero tal redefinición carecería de objeto, porque con ella dejaríamos fuera de la redefinición, precisamente, aquello que tratamos de redefinir. Así pues, según Searle, la irreductibilidad de la conciencia es "una consecuencia trivial de la pragmática de nuestras prácticas definitorias" (Ibíd.: 122 del original, 132 de la traducción): no podemos reducir ontológicamente la conciencia porque las reducciones ontológicas consisten en eliminar las apariencias subjetivas y quedarnos con lo que nos interesa de un determinado fenómeno tras comprobar que, dada su verdadera historia causal, dichas apariencias se deben a determinada estructura causal y no son nada más que determinados efectos de la misma. Esto no podría aplicarse a la conciencia dado que en su caso la apariencia subjetiva es la realidad misma del fenómeno y dado que en el caso de la conciencia lo que nos interesa es, precisamente, la propia apariencia. En este sentido, cabe entender que, en la propuesta de Searle, "si un fenómeno puede o no ser reducido (…) depende de nuestros intereses" (Nida-Rümelin, 2002: 218τ).

La conciencia sería pues reductible e irreductible: podemos redefinirla en términos de su substrato causal, aunque ello carecería de sentido, por cuanto la reducción ontológica del calor a movimiento cinético deja sin reducir la experiencia subjetiva del calor del mismo modo que la reducción de la conciencia a fenómenos neurofisiológicos dejaría sin reducir la experiencia consciente. El motivo por el cual la conciencia es ontológicamente irreductible según Searle, repitámoslo, es que el patrón habitual de reducción está destinado a dejar fuera aspectos perceptivos o aparienciales de los fenómenos reducidos y tratarlos así en términos de su estructura física subyacente. Consistiendo la conciencia en su apariencia, esa distinción entre apariencia y realidad no podría trazarse y quedaría así excluida de dicho patrón habitual de reducción.

El argumento de Searle conduce a una serie de interrogantes. En primer lugar, parecía que Searle venía defendiendo que la conciencia se caracteriza por esa ontología de primera persona en la que tanto hincapié hace. Mas, ¿hace con su argumento depender dicha ontología de la *pragmática de nuestras prácticas definitorias*? Searle repite una y otra vez que lo crucial de la conciencia es su ontología de primera persona y que esa ontología es irreductible e ineliminable. Leer ese tipo de frase en sus textos una y otra vez hace que resulte sorprendente encontrar que cuando por fin Searle argumenta explícitamente acerca de los motivos por los cuales la conciencia es irreductible diga sin ambages que ello no es sino una consecuencia trivial de la pragmática de nuestras prácticas definitorias. ¿Que la conciencia sea irreductible resulta no de esa especialísima ontología suya en la que Searle insiste sino de nuestras prácticas definitorias? Parece que después de escribir en decenas de páginas expresiones como "ineliminable ontología de primera persona" o "irreductible ontología subjetiva de la conciencia", Searle sólo se refería a determinado aspecto de nuestra economía léxica y no a determinado fenómeno existente en el mundo, no a determinado fenómeno del que parecía venir deseando convencernos de que es enteramente real y diferente en un importante sentido de la mera actividad neurofisiológica: en el sentido de que su ontología es ineliminablemente subjetiva. Muchos habíamos llegado a creer que Searle trataba de argumentar en favor del carácter ontológicamente distintivo de la conciencia y en favor de su irreductibilidad a lo físico dado, precisamente, dicho carácter distintivo. Pero, curiosamente, parece que al final la irreductibilidad de la conciencia no tiene demasiado que ver con la ontología y se limita a ser una consecuencia trivial de nuestras prácticas definitorias. Searle da nuevamente la impresión de moverse entre dos tierras: ¿es la irreductibilidad ontológica de la conciencia una consecuencia de su ontología ineliminablemente subjetiva –como parecía venir sugiriendo Searle– o lo es de nuestras prácticas definitorias? Diez años después de proponer el señalado argumento, y para tranquilidad de todos los que con él pensamos que llevábamos tiempo malinterpretando a Searle, éste, en su estilo habitual, respondió a nuestra última pregunta apuntando que la ontología de primera persona de la conciencia no puede ser reducida, sencillamente, porque es una ontología de primera persona, con lo cual dejamos de tener un argumento a favor de la irreductibilidad, pero sabemos al menos que ella deriva de la ontología de la conciencia antes que de nuestras prácticas o intereses.

> What is the difference between consciousness and other phenomena that undergo an ontological reduction on the basis of a causal reduction, phenomena such as colour and solidity? The difference is that consciousness has a first-person ontology; that is, it only exists *as experienced* by some human or animal, and therefore, it cannot be reduced to something that has a third-person ontology, something that exists independently of experiences. It is as simple as that. (Searle, 2002b: 60-61).[77]

En segundo lugar: ¿Ha de obedecer toda reducción fisicalista a lo que Searle denomina patrón habitual y define refiriéndose a la necesidad de dejar de lado las apariencias? Paul Churchland (Churchland, 1996) ha puesto en duda que así sea y ha ofrecido argumentos destinados a mostrar que un hecho histórico como el de que las reducciones fisicalistas hayan venido dejando fuera las apariencias o efectos de los hechos reducidos sobre la conciencia humana no significa que el patrón haya de repetirse indefinidamente, ni que dicho patrón tenga ningún significado ontológico: lo único que desvelaría sería nuestra provisoria ignorancia científica, es decir, un hecho circunstancial antes que una *necesaria* (recordemos que Searle entiende que la conciencia es irreductible dada la *definición* de reducción que él maneja; dicha *necesidad* parece encajar mal, por otra parte, en el esquema searleano: contra su realismo ingenuo, *nuestra* epistemología estaría determinando *la* ontología) división en la naturaleza que haría de todo fenómeno físico, a excepción de la conciencia, un hecho ontológicamente reductible.

En tercer lugar: ¿cuál hemos de entender que es la diferencia entre ambos tipos de reducción? La duda acerca de si existe una verdadera diferencia entre lo que Searle ha denominado reducción causal y lo que ha denominado reducción ontológica ha sido sucinta y contundentemente articulada por Paul Churchland en el artículo de 1996 al que acabamos de referirnos, en el que plantea 1) que Searle asimila erróneamente reducción ontológica y eliminación ontológica,[78] y 2) que, sencillamente, no existe algo a lo que denominar "reducción causal", que se trata de una categoría vacía o equivalente, en cualquier caso, a la de "reducción ontológica": "there simply is no further category

[77] Cursivas en el original. En la misma línea, Searle (2006: 105) ha señalado que la ontología de primera persona de la conciencia implica que sólo existe en tanto experienciada y que *por tanto* no puede reducirse a nada que posea una ontología de tercera persona. El núcleo de su propuesta antirreduccionista vuelve a situarse en la ontología de la primera persona antes que en nuestras prácticas lingüísticas. Un año después insistirá en que la imposibilidad de la reducción se debe a la ontología de la conciencia en términos más explícitos si cabe: "Because consciousness has a first-person ontology, it cannot be reduced" (Searle, 2007b: 171); "the causal reduction does not lead to an ontological reduction because consciousness has a first-person or subjective ontology, and for that reason cannot be reduced to something that has a third-person or objective ontology" (Ibíd.: 175).

[78] Searle, en cierto sentido, se guardará posteriormente las espaldas indicando que, por lo que a la conciencia respecta, la distinción entre reducciones no-eliminativas y reducciones eliminativas no puede trazarse, porque ambas confluirían, en último término, en la negación de las características definitorias de la conciencia (Searle, 2007b: 171). Quede ajuicio del lector la efectividad de semejante estrategia.

or 'half-way house' –Searle's so-called 'causal reduction'– distinct from ontological reduction" (Churchland, 1996: 224). Y, ciertamente, cabría preguntar, ¿qué hemos de entender que es la reducción ontológica del calor una vez realizada su reducción causal? ¿Hay dos etapas definidas y diferenciables antes de llegar al término de una reducción ontológica? Igualmente, y pasando a considerar eliminaciones ontológicas –pues, como Churchland señala, Searle las confunde con las reducciones ontológicas–, conforme fue alcanzándose una comprensión de la neurotransmisión en términos fisiológicos y químicos modernos, fue abandonándose la noción de aquellos espíritus animales de la medicina galénica,[79] pero, por lo que toca a las –eliminaciones o– reducciones (de espíritus a neurotransmisores, primeros mensajeros, factores de transcripción, etc.): ¿dónde empieza la ontológica y acaba la causal? ¿Pueden distinguirse las dos reducciones que necesita Searle? ¿Se redujeron primero causalmente los espíritus para luego ser eliminados de nuestra ontología?

Por otra parte, si hay algo que reducir ontológicamente después de realizar una adecuada reducción causal, máxime cuando no puede realizarse este movimiento (como en el caso de la conciencia, según Searle), parece que sólo cabe educir una consecuencia (ya implícita en anteriores apreciaciones): la de que hay una determinada parte de nuestra ontología que nada tiene de causal.

Una última dificultad ontológica relacionada con el tratamiento de la causalidad en la filosofía de la mente de Searle tiene que ver con su intención de mantener la tesis de la clausura causal del mundo físico y defender al mismo tiempo la eficacia causal de lo mental (vid., v. g., Searle 2000b; 2001: 41). A pesar de haber sido éste lugar para interesantes polémicas (Kim, 1995; 1998: cap. 2; Searle, 2003), y a pesar de constituir el problema de la causación mental, de por sí, un núcleo de gran amplitud dentro del debate contemporáneo en filosofía de la mente (hasta el punto en que algunos lo entienden como el núcleo del problema mente-cerebro: vid., v. g., Dardis, 2008) nos referiremos aquí casi de pasada a las dificultades que el señalado problema trae consigo para el planteamiento de Searle. Sin entrar en detalles, las dos tesis citadas (clausura causal del mundo físico y eficacia causal de lo mental) conducen al conjugarlas a problemas como el de la sobredeterminación causal. Searle cree encontrarse en una buena situación para

[79] Cuya andadura en la medicina occidental se prolongara durante milenio y medio, adoptando, como es lógico, diferentes formas. Es famoso el tratamiento que Descartes hiciera de los mismos en *Les passions de l'âme* o en *L'Homme*, como lo es la crítica realizada en ese mismo siglo, el XVII, por Harvey a la noción de esos espíritus animales. En nuestro país, el médico y humanista vallisoletano Gómez Pereira los definió en el siglo anterior en términos de *phantasmata*, como "ciertos corpúsculos inmateriales influidos de una forma oculta por los objetos externos" (Pereira, 1554: 20).

evitar tal problema, y alega que el mismo no surge para él, dado que uno y el mismo sistema admite *descripciones causales* a diferentes niveles (Searle, 1995b: 219). Searle defiende incluso la existencia de un tipo vertical descendente de causación, de lo mental a lo fisiológico, y entiende que su propuesta no resulta problemática porque las propiedades mentales del macronivel, a pesar de ser *causalmente reales*, efectivas, se hallan *realizadas* en los elementos del micronivel (Ibíd.).[80] Surgen con esto dos problemas. El primero puede resumirse en dos frases: para que las relaciones de causalidad de las que habla Searle –bien sean ascendentes o descendentes– resulten creíbles, primero debiera resultarlo la idea de una forma monádica y sincrónica de causalidad acaecida dentro de los márgenes de una y la misma entidad, una forma de causalidad que operaría, leyendo a Searle al pie de la letra, ¡entre dos descripciones diferentes de uno y el mismo fenómeno! ¿Puede una descripción satírica de Eric Arthur Blair actuar causalmente sobre una descripción dramática de George Orwell? El segundo, por su parte, consistiría en que, como veíamos poco más arriba, es difícil elucidar hasta qué punto esas descripciones de diferentes niveles causales responderían a nuestras prácticas descriptivas antes que a la efectiva existencia de dos niveles diferentes y reales de causación –elucidación que resulta incluso más acuciante y menos prometedora cuando Searle afirma: "consciousness has no causal powers of its own in addition to the causal powers of the underlying neurobiology" (Searle, 2002b: 60)–. ¿Son pues reales y diferentes de los del micronivel los poderes causales del macronivel o pueden ser entendidos como nada más que una forma abreviada de referirnos a procesos causales que pueden, en último término, explicarse enteramente sin abandonar el micronivel?[81]

[80] Es interesante hacer notar en este punto que esta idea de una causación vertical descendente de lo mental a lo fisiológico evidencia que la concepción de la causalidad elaborada por Searele comporta un dualismo ontológico que a estas alturas debería resultar ya enteramente evidente: esta propuesta causal, que convierte el cerebro al completo en el centro de interacción entre lo mental y lo físico que Descartes ubicara en la glándula pineal, presenta a la ontología irreductiblemente subjetiva de lo mental como algo heterogéneo respecto del cerebro que, si hemos de hacer caso a Searle, la causa y "realiza" (venga esto a significar lo que quiera que venga a significar), pues con la misma nos hallamos ante una interacción causal entre una determinada ejemplificación de un nivel ontológico superior, la mente de Juan, sobre otra de uno inferior, su cerebro. La apelación a diferentes niveles descriptivos de una y la misma cosa no parece zanjar el asunto (a no ser que *causa sui* resulte finalmente una noción enteramente clara), porque "el naturalismo biológico [sigue] toma[ndo] lo mental como algo ontológicamente diferenciable de aquello que lo causa" (Pérez Chico, 1999: 135).

[81] Otra cuestión que cabría discutir dentro del marco del tratamiento de la causalidad en el contexto de la ontología de lo mental propuesta por Searle –y particularmente en el contexto de su concepción de la causación mental– pero en la que no podremos detenernos aquí tendría que ver con el hecho de que Searle parece obviar la complejidad de los marcos teóricos destinados a dar cuenta de la causalidad neurofisiológica. Retomando la discusión acerca de la causación mental desde la perspectiva de estos marcos teóricos, cabría confrontar el modelo de causación expuesto en el diagrama presentado en Searle (2004a: 210) con perspectivas de grano más fino acerca del problema de la causación mental. El aludido diagrama vendría a resumir la propuesta searleana: un evento neurofisiológico causa el siguiente, pero los eventos mentales

Relacionadas con las dificultades a las que se enfrenta la ontología de Searle se hallan asimismo algunas dificultades epistemológicas cuyo carácter se encuentra ya implícito en nuestra denuncia del naturalismo biológico como dualismo biológico. Nos referiremos brevemente a la que consideramos la más acuciante de entre las mismas. Searle distingue entre objetividad epistemológica y objetividad ontológica para plantear 1) que no toda la realidad es ontológicamente objetiva, porque no podemos negar la existencia de la conciencia como fenómeno ontológicamente subjetivo, y 2) que una ciencia epistemológicamente objetiva de la conciencia (en tanto fenómeno ontológicamente subjetivo) es, en cualquier caso, posible. El principal problema es aquí el de la vaguedad de la propuesta de Searle, que se queda en el plano nominal y no ofrece en ningún caso siquiera pistas acerca del modo de integrar en los modelos explicativos disponibles en las disciplinas científicas pertinentes un fenómeno radicalmente diverso del resto de los tratados por cualquiera de las mismas. Searle se inclina por un tratamiento científico de la conciencia desde la biología, particularmente desde las neurociencias, pero tampoco aquí ofrece ninguna clase de perspectiva o pista en el sentido señalado.[82] Se limita a proponer que una ciencia epistemológicamente objetiva de la conciencia es posible,[83] y que la pasividad de objetividad epistemológica de la conciencia no puede ponerse en duda, dado que existirán, en cualquier caso, afirmaciones cier-

causados por y realizados en dichos eventos neurofisiológicos no dejan de ser tan causalmente eficaces como aquellos estados neurofisiológicos que los causan y en los que se realizan. Este planteamiento ha sido criticado desde perspectivas cuánticas y no lineales. El modo en que los eventos y procesos neurofisiológicos son en el planteamiento searleano tenidos en cuenta excluye la influencia que los fenómenos cuánticos pudieran tener al seguir las flechas del diagrama, esto es, por ejemplo, al pasar de un estado fisiológico al siguiente. En este sentido, Henry Stapp, en su crítica de la concepción searleana de la causalidad, ha afirmado: "the dictates of classical physics cannot be relied upon to reveal the whole truth in the case of brain activities. This is because brain dynamics depends critically upon such things as flows of ions into nerve terminals, and the dynamical properties of ions are not correctly specified by classical physics" (Stapp, 2011: 143). De forma análoga, y sin recurrir a la física cuántica sino a una concepción del funcionamiento del cerebro basada en la no-linealidad, la autoorganización y el comportamiento caótico, Freeman & Skarda (1991) han argumentado, por lo que toca a la línea del diagrama según la cual un evento fisiológico causa el siguiente, que hablar de la causa concreta de un evento en el cerebro resulta arriesgado, cuando no imposible, cosa que a nadie debiera extrañar habida cuenta del modo en que el azar ha venido insertándose en nuestra concepción del universo a todos los niveles –en este sentido, el Nobel de Química Ilya Prigogine, partiendo de la consideración de sistemas dinámicos y de la física del no-equilibrio, ha insistido en que el azar, la probabilidad, la indeterminación, la historicidad y la irreversibilidad, antes que tener que ver con nuestra perspectiva o con la insuficiencia de nuestras mediciones o recolecciones de datos, pertenecen esencialmente al universo a todos los niveles: "el azar permanece esencial" (Prigogine, 1988: 86 de la traducción).

[82] Lo cual, desde dentro de la concepción searleana, habría de ser tenido por una grave falencia, dado que es el propio Searle quien entiende que "[o]ne of the tasks of the philosopher is to get the problem into such a shape that it can be subject to experimental testing in neurobiology" (Searle, 2006: 102).

[83] "You can have a perfectly objective science of an ontologically subjective domain" (Searle, 2007b: 172).

tas y afirmaciones falsas acerca de experiencias conscientes: así, cuando digo que veo una postimagen mientras en efecto la veo, expreso un juicio epistémicamente objetivo, pues no depende de intereses, gustos o preferencias personales. El camino searleano hacia la objetividad epistémica en el tratamiento científico de la conciencia acaba aquí. No ofrece ninguna otra indicación[84] y, ciertamente, cabe dudar que con lo dicho logremos un avance significativo. Es más que probable que Searle esté al corriente de que la objetividad epistémica que requiere el conocimiento científico no se detiene en la constatación de hechos en base a criterios epistémicamente objetivos desde el punto de vista de su contenido aunque epistémicamente subjetivos desde el punto de vista de su acceso, sino que requiere, por el contrario, de la búsqueda de criterios públicos no sólo para la constatación de la existencia (cosa que ya cabe dudar que alcance a ofrecer el marco teórico de Searle), sino, por ejemplo, para la manipulación experimental unívoca y sistemática de los fenómenos a estudiar con independencia de puntos de vista, gustos, sentimientos, preferencias, etc. Searle no nos dice cómo dar el salto que haga de la subjetividad ontológica objeto pasible de un tratamiento epistémicamente objetivo, y, de hecho, nos invita a buscar la explicación de la conciencia en rasgos del cerebro que, en tanto rasgos físicos, pueden concebirse y describirse a la vez –según Searle, en función del nivel descriptivo escogido (vid., v. g., Searle 2007b: 176)– como rasgos ontológicamente objetivos y ontológicamente subjetivos.[85] Lo que sí nos dice Searle es que dichos rasgos, cuando los consideramos en su ontología subjetiva, traen consigo una relación epistémica privativa: sólo el dueño del cerebro que los porte tiene acceso a los mismos. Esto, ciertamente, hace que entre dichos rasgos del cerebro y el resto de los fenómenos estudiados por cualquier disciplina biológica medie un trecho considerable que Searle no explica cómo recorrer: se limita a presentar dichos rasgos del cerebro como rasgos biológicos *enteramente ordinarios* (vid. v. g., Searle, 2002a: 2). En la misma línea, la segunda dificultad epistemológica que, en el artículo con el que abríamos esta sección, Guerrero del Amo detecta y critica en el planteamiento de Serle tiene que ver con la necesidad a que su propuesta aboca de reformular el marco científico actual a fin de que éste logre dar cabida a y cuenta de la conciencia como un fenómeno ontológicamente subjetivo. El problema sería en este caso, según Guerrero del Amo, el de afrontar dicha

[84] Quizá debamos contar aquí como excepción su preferencia por un abordaje neurobiológico holístico ("unified field" conception of consciousness) antes que uno dividido en bloques ("building block" model of consciousness) (Faigenbaum, 2003: 51; Searle, 2000a: 53 y ss.; 2007b: 175).
[85] "Where consciousness is concerned, there are first-person phenomena and third-person phenomena" (Searle, 2007b: 177).

reformulación partiendo del supuesto searleano según el cual la conciencia es "ineliminablemente subjetiva y no puede convertirse en objeto de observación del mismo modo en que pueden los objetos y estados de cosas objetivamente existentes" (Searle, 1992: 98 del original, 109 de la traducción). Como vemos, se trata del mismo tipo de crítica epistemológica que hallábamos ya implícita en nuestra denuncia del naturalismo biológico como dualismo biológico. ¿Cómo afrontar pues la señalada reformulación en vista de esta imposibilidad de equiparar la conciencia al resto de objetos de las ciencias? La respuesta de Searle a tal interrogante ha venido consistiendo (Searle 1992; 1997a) en señalar que, a pesar de que nuestro acceso sea distinto en el caso de los estados mentales conscientes y en el de los hechos u objetos con una ontología de tercera persona, podríamos hacer uso de medios indirectos y correlacionar conducta y neurofisiología para integrar la conciencia en un marco científico adecuado. El problema para Searle es que la conciencia, con la ontología de primera persona en la que el de Denver incide, sigue al parecer quedando lejos del alcance de semejante propuesta, que parece más bien aproximarse a la heterofenomenología dennettiana. Guerrero del Amo subraya que la objetividad epistemológica de Searle, para suministrar la posibilidad de la integración de la conciencia en un marco explicativo científico, no sólo debería implicar que la verdad de un enunciado (como "me duele la espalda") sea independiente de actitudes sentimientos o puntos de vista personales, sino que, además, debería presentarse de la mano de la posibilidad de acceder al contenido de dicho enunciado de forma análoga por diferentes observadores, cosa difícilmente exigible teniendo en cuenta la subjetividad ontológica de los estados mentales conscientes en la que Searle insiste (Guerrero del Amo, 2001: 311). El problema de Searle, tal y como Guerrero del Amo lo presenta en el artículo que venimos comentando, sería el de mantener la subjetividad ontológica de la conciencia, es decir, preservar la subjetividad como un componente irreductible de la realidad, y al mismo tiempo defender el modelo de objetividad proporcionado por las ciencias físicas y biológicas. En este punto, Guerrero del Amo cita (dislocando, por así decir, la cita respecto de su contexto) José Luis Pinillos para poner de relieve que el drama de Searle consiste en "pretender una ciencia objetiva de una realidad subjetiva, sin poseer un concepto de ciencia adecuado para tal empresa" (Pinillos, 1978: 163). Nuevamente, la propuesta de Searle se ve acosada por tensiones que desembocan en divisiones conflictivas, divisiones que dibujan los contornos de su planteamiento sin borrar de los mismos el perfil de un dualismo conceptual que se muestra incapaz de abandonar.

El objetivo de Searle consistía en abandonar el aparato conceptual heredado y abrir una tercera vía entre dualismo y materialismo. En vista de los numerosos impasses a los que su naturalismo biológico conduce, algunos han llegado a afirmar que aquél al que su prosa ha venido sirviendo sería más bien el de la elaboración de una "ineficaz retórica interminablemente repetitiva" (Colomina Almiñana, 2010: 174).

Es muy probable que Searle respondiera a todas las dudas que su propuesta nos ha suscitado acusándonos de dualistas conceptuales. Si no logramos ver el sentido en que su propuesta ha superado cualquier tipo de dualismo, alegaría, ello se debe a nuestra incapacidad para dejar atrás un aparato conceptual cuyo núcleo problemático encuentra Searle, principalmente, en la asunción, bien implícita o explícita, de una concepción según la cual lo físico y lo mental resultan, necesariamente, mutuamente excluyentes. No obstante, baste señalar, con Garrett (1995: 209-210), que el comodín del dualismo conceptual, que Searle utiliza para despachar todas las discusiones acerca del problema de la conciencia presentando todas y cada una de las posturas vigentes y abandonadas como incoherentes, no es explicitado por Searle sino de un modo fortuito y superficial (cosa curiosa habida cuenta del portentoso partido que saca del mismo), no entrando el de Denver a precisar el modo en que el dualismo conceptual causa los estragos que, haciendo caso de su propuesta, hemos de entender que causa en cada punto del debate contemporáneo. Quede a juicio del lector la decisión acerca de si dicha acusación lograría disipar las dudas que compendiosamente hemos planteado.

# CAPÍTULO 11

## LA “CONCIENCIA” “EXPLICADA” POR DENNETT

### 1._El rompecabezas de Dennett: ciencia, objetividad y conciencia

Desde su irrupción en la escena filosófica a finales de los años sesenta la producción de Daniel Dennett se ha caracterizado por una solícita atención a las áreas científicas pertinentes, es decir, las vinculadas con los extremos de los que ha venido ocupándose. De este modo, como indicábamos en el primer capítulo de la segunda parte, ya en su tesis doctoral ofreció un tratamiento científicamente informado de las discusiones abiertas en la filosofía de la mente del momento. La conciencia es hoy una de las centrales cuestiones en liza en filosofía de la mente. Sin embargo, una discusión filosófica científicamente informada en este punto no resultaba en el momento en que preparaba su tesis doctoral tan sencilla como pudiera resultarlo una discusión de tales características en filosofía de la física. ¿Por qué? Sencillamente porque, como Dennett (1978) constatara transcurrida una década desde su primera publicación, no era aquél un momento en que los científicos estuvieran prestando demasiada atención al problema de la conciencia, sino más bien todo lo contrario.[1] ¿Y a qué podríamos entender que se debía aquella incuriosa tendencia para con la conciencia y su estudio científico? Cabría tratar de responder a esta pregunta haciendo notar que, desde sus orígenes en el siglo XVII, la ciencia moderna ha propiciado –y se ha visto propiciada por– un proceso de acendramiento mediante el cual *el ojo pretendía alcanzar a ver desde ninguna parte* (Nagel, 1986) expurgando de subjetividad

[1] Entre el momento en que Dennett redacta este artículo pionero y el momento en que Baars formula su Teoría del Espacio de Trabajo Global transcurriría una década, y cerca de dos serían necesarias para que la neurobiología estallara en un abigarrado marco de propuestas teóricas y metodológicas para el estudio de las bases neuronales de la conciencia.

el conocimiento científico, idealmente dirigido a la neutralidad objetiva independiente de puntos de vista particulares y sesgos personales. Pero, ¿no es precisamente la conciencia esencialmente –*ontológicamente*, diría Searle– subjetiva, no se trata de algo que pertenece indefectiblemente a un individuo, que le es privada, que resulta privativa y exclusiva, que está necesariamente dotada de un punto de vista particular, de una peculiar perspectiva? Parece indudable: la ciencia no estudia nada similar.[2] Con todo, el proyecto de Dennett es el de hacer encajar las piezas (objetividad y conciencia) de tan desconcertante rompecabezas situándolas en el mismo plano. Así, opta por un camino que algunos retendrían en el cedazo "pues peor para los datos"; es decir, no resulta inusual que se caracterice la estrategia propuesta por Dennett para el estudio de la conciencia como sigue: si la ciencia no se adapta al objeto, adaptemos pues el objeto a los estándares científicos. Y, en efecto, si definimos la conciencia como un fenómeno mental subjetivo, interno, privado, perspectivo y no susceptible de ser compartido,[3] es probable que nos sintamos engañados cuando en el breve capítulo introductorio a *The Intentional Stance* (Dennett, 1987) o en las primeras páginas de *Consciousness Explained* (Dennett, 1991a) –ya en el preludio y al comienzo de la subsiguiente primera parte– se nos ponga al corriente de que lo que se dispone a ensayar el autor es –respectivamente– una aproximación a la mente en tercera persona y una explicación de la conciencia conforme a los tradicionales y objetivos supuestos de trabajo propios del método científico. De este modo, no son pocos los filósofos que, desde mediados de los setenta y abanderados por Nagel, entienden que desde una perspectiva objetivista como la de Dennett, una perspectiva que pretenda, por así decir, exhibir en público lo que de hecho sólo se muestra –se experimenta– privadamente, se estarían dejando de lado, precisamente, los aspectos medulares y característicos de la conciencia, sus rasgos distintivos –es decir, precisamente, los datos que la teoría debería explicar–, motivo por el cual se ha sostenido con frecuencia que, "sencillamente,

[2] Precisamente el hecho de que los estados fenoménicamente conscientes sean *privados* y estén dotados de *perspectiva* –"subjetividad perspectiva"– son los dos primeros problemas filosóficos de la conciencia analizados por Tye (2007), los dos primeros que causan este tipo de perplejidad: "ningún otro fenómeno del que quepa ofrecer explicación científica parece poseer semejantes propiedades".

[3] Estos son los primeros rasgos de la descripción de la conciencia que realizan Edelman y Tononi al final de la primera parte de *A Universe of Consciousness*, pero advierten: una explicación científica de la conciencia no puede sustituir a una experiencia consciente –y, por nuestra parte, consideramos oportuno añadir que ninguna explicación científica equivale a ningún hecho o fenómeno, pues no es éste su cometido, sino más bien el de hallar las condiciones en las que nos quepa anticipar: dadas estas circunstancias es imposible, posible o "necesario" que suceda tal o cual cosa de tal o cual modo y por tales y cuales motivos.

excluye *a priori* lo que hay que explicar y, por tanto, no ofrece ninguna explicación de la conciencia" (Saéz Rueda: 392).[4]

Hacíamos mención de una serie de características de la conciencia que pueden ser tenidas por definitorias y que serían aquéllas a las que los señalados filósofos se atendrían en su crítica al objetivismo dennettiano. Retengamos ahora dos de especial interés en relación con el planteamiento en tercera persona propuesto por Dennett: la privacidad y la perspectiva. Parece intuitivamente cierto que las ciencias de la mente-cerebro habrían de dar poco menos que un salto mortal para que pudiéramos representarnos la experiencia consciente como desligada de dichas características. Un estado consciente es siempre el estado consciente de un organismo, y le es pues privado no sólo en este sentido, sino también en cuanto tal organismo, digamos, se halla con tal estado en una relación radicalmente diferente –es decir, siempre diversa– a la que desde el punto de vista externo de la tercera persona pueda mantener cualquier otro organismo con tal estado. Precisamente el hecho de que el organismo se halle relacionado con sus estados conscientes de este modo privativo (nadie puede entrar como él en dicha relación) implica la segunda de las características señaladas: sólo a él le es dable decir "tal y como a mí me parece…" (independientemente de lo expresivo que sea el contenido de la enunciación que venga a sustituir a los puntos suspensivos; en otras palabras, con independencia de la capacidad del sujeto para plasmar su experiencia en su discurso y su conducta), es decir, sólo él se halla situado en la particular perspectiva proporcionada por su experiencia consciente. Estas dos características pueden ilustrarse con ejemplos tomados del lenguaje ordinario. Así, la expresión "meterse en la cabeza de alguien" puede no sólo pretender indicar "comprender los motivos e intenciones de ese alguien", sino también "sentir sus estados anímicos o percibir los estímulos ambientales del modo en que él los experimenta", acceder a ese ámbito privado en y desde el que los siente y percibe –algo imposible si en verdad la experiencia consciente se caracterizada, como parece, por la privacidad y la perspectiva.

Dennett, como señalábamos en nuestra referencia a los primeros compases de *The Intentional Stance* y las primeras páginas de *Consciousness Explained*, no suele –contra la acusación de Searle (1997a: 100-101 del original, 96 de la traducción)– ocultar sus

[4] Puede resultar pertinente en este ínterin apuntar a la dudosa ecuanimidad con la que el autor del texto citado alude repetidamente (Sáez Rueda: 385 y, en un contexto particularmente interesante, 389) a la heterofenomenología dennettiana (que introducimos más abajo) como una "afrenta".

intenciones y advierte desde el principio algo así como: "¿Ésos te parecen los rasgos 'esenciales'? ¡Muy bien! Tratemos de ver qué pasa si adoptamos respecto del fenómeno en cuestión un punto de vista objetivo, nos atenemos a los datos y tratamos de sacar las conclusiones pertinentes". ¿Cómo adaptar un fenómeno en el indicado sentido privado y perspectivo a los estándares de neutralidad y publicidad del método científico? ¿Se adecua a su objeto un método para el estudio de la conciencia que adopte sobre ella la perspectiva de la tercera persona? Éste es el reto que una teoría como la de Dennett asume:

> El desafío reside precisamente en construir una teoría de los eventos mentales, usando los datos permitidos por el método científico. Tal teoría sólo podrá construirse a partir del punto de vista de la tercera persona, porque toda ciencia se construye desde esta perspectiva (Dennett, 1991a: 71 del original, 84 de la traducción).

Desde el punto de vista de los críticos de este objetivismo, la necesidad en que el método científico se halla de objetividad, publicidad y cuantificabilidad a expensas de subjetividad, privacidad y perspectiva, excluye la posibilidad de ofrecer desde el mismo un tratamiento que haga justicia a las características definitorias de la conciencia, pues desde una perspectiva en tercera persona como la dennettiana tales características habrían, necesariamente, de ser obviadas. Sin embargo, puede que tanto Dennett como sus críticos hayan pasado por alto que la objetividad que busca el método científico no es otra cosa que la replicabilidad, la posibilidad de alcanzar unos determinados resultados mediante unos determinados procedimientos.

Más allá del modo en Nagel y otros han defendido que la experiencia consciente permanecerá fuera del alcance de la investigación científica, una de las consecuencias antiintuitivas de una posición objetivista como la ensayada por Dennett es que, de cara a tratar la conciencia desde un punto de vista científico, el de Boston prescribe como punto de partida una perspectiva neutral acerca de la misma. Esto viene a significar: nada de claridad y transparencia, incorregibilidad o inmunidad introspectiva al error. Dicha neutralidad como punto de partida pretende advertir al investigador de la conciencia de que sabemos tanto sobre ella como sobre los volcanes, las reacciones de oxidación o la herencia de caracteres biológicos antes de estudiar científicamente tales fenómenos, de modo que debemos permanecer neutrales respecto de lo que la conciencia *sea*; es decir, debemos abstenernos de pronunciarnos a este respecto, a pesar de la tradición y el sólido círculo de intuiciones que se cierne en torno a nosotros persuadiéndonos de que la con-

ciencia es de hecho algo que conocemos directa y claramente.[5] Bien es cierto que la conciencia es de momento –y comparada con nociones como animal, planeta, haploide o deriva continental– una noción vaga y amplia, dentro de la cual no sólo caen sin duda diversas subclases, sino que, además, por lo pronto no existe un acuerdo general acerca de cuántas y cuáles serían esas subclases; además, y por si fuera poco, las fronteras entre éstas subclases o formas de conciencia pueden resultar difusas. Lo antedicho trasluce o pretende, en definitiva, traslucir que reina actualmente la discusión conceptual al tiempo que una pluralidad de enfoques teórico-experimentales difícilmente reducible a unos cuantos puntos en común. Así, no contamos con una definición unánimemente aceptada de conciencia, ni, por tanto, con una noción compartida y a todas luces evidente o un concepto diáfano de la misma. En este sentido ha afirmado Johnson-Laird que nadie sabe qué es la conciencia, añadiendo que desconocemos asimismo qué es lo que hace y cuál es su función (Johnson-Laird, 1983a: 448). No obstante, cabe conceder este punto, el del desconocimiento científico y la vaguedad conceptual –una situación que, según Bunge y Ardila, "recuerda la de la física en los primeros días de la electricidad y el magnetismo" (Bunge & Ardila: 1987: 262 de la traducción)–, sin comprometerse por ello con la necesidad de una radical neutralidad como punto de partida, porque la misma resulta de hecho imposible. El mero hecho de haber llevado una vida consciente orientará desde el comienzo cualquier teoría de la conciencia que alcancemos a desarrollar. Algunas teorías epistemológicas (generalmente relativistas, como las de Hanson o Kuhn, pero también antipositivistas, como el holismo confirmacional Duhem-Quine) ponen en la pista de la imposibilidad de una total neutralidad de nuestras teorías, y parece que aquí –es decir, en lo que a una teoría de la conciencia respecta– resultan más pertinentes y plausibles que en cualquier otro lugar.

Dennett intenta ofrecer un método neutral para el estudio de la conciencia, un método neutral que comienza por resultar un tanto parcial: sólo las criaturas lingüísticas son susceptibles de ser estudiadas mediante dicho método. No obstante, el tipo de conciencia que dicho método está destinado a estudiar es la humana, de modo que podemos pasar eventualmente por alto el carácter parcial de semejante método neutral. Dicho método es la heterofenomenología, y mediante él pretende Dennett salvar el –*¿aparente*?– abismo

[5] Sin la intención de entrar en detalles nos gustaría señalar que tal vez resulte provechoso confrontar esta neutralidad con las implicaciones metodológicas y epistemológicas de propuestas tales como la de la carga teórica de los hechos de Hanson –también conocida como problema de la infradeterminación de la teoría por los datos: "*hay más en la vista de lo que se encuentra en el ojo*" (Hanson, 1958: 34τ)– o la tesis del holismo confirmacional de Duhem-Quine.

que separa el *what it is like* de la pertinente objetividad propia de la investigación científica; con él pretende, pues, conciliar investigación científica objetiva y experiencia consciente subjetiva. Detengámonos en dicho método unos instantes antes de pasar a ocuparnos brevemente del modelo de la conciencia que Dennett elabora y defiende.

## 2._Un método para la fenomenología

Consideremos brevemente en primer lugar los motivos por los que Dennett rechaza la fenomenología[6] y propone sustituirla por su heterofenomenología como método neutral y objetivo para el estudio (eminentemente descriptivo) de la conciencia. Dennett entiende que lo que denomina *presunción de la primera persona del plural* –dar por sentado que la descripción de lo que el fenomenólogo encuentra en su propia experiencia es extensible a la experiencia de los demás– no resulta justificable –sino que más bien es una traicionera fuente de errores–, y menos a la luz de nuestra falibilidad introspectiva, la cual ilustra Dennett a lo largo de *Consciousness Explained* con diversos ejemplos destinados a probar que, por así decir, en la introspección no se llevan a cabo hialinos actos perceptivos análogos a los de la percepción externa: no nos asomamos y constatamos, sino que en gran medida teorizamos inadvertidamente acerca de los contenidos de nuestros mundos interiores. En esta situación, el estudio científico de la conciencia requerirá de un método capaz –en principio– de separar la paja del trigo y soslayar tales óbices. ¿Cuál es, según Dennett, tal método?

Partimos de la asunción de que cualquier investigación empírica de la conciencia requerirá del uso del lenguaje, con lo cual unos cuantos supuestos han de añadirse al de partida: que los sujetos experimentales entienden las instrucciones de los investigadores, que su conducta lingüística es significativa y que el significado de la misma tal y como el investigador acabará por interpretarla responde al significado pretendido por el sujeto para ella. Aceptados estos supuestos, la primera etapa de un estudio heterofenomenológico consistirá en el registro de la conducta lingüística del sujeto y los experimentadores. Este registro está destinado a conformar un texto heterofenomenológico que los investi-

[6] "I studied Husserl and the other Phenomenologists with Dag Føllesdal at Harvard as an undergraduate, and learned a lot. My career-long concentration on intentionality had its beginnings as much with Husserl as with Quine. But part of what I thought I learned from those early encounters is that reading the self-styled Husserlians was largely a waste of time; they were deeply into obscurantism for its own sake. I may have picked this attitude up from my graduate advisor, Gilbert Ryle, who was himself a masterful scholar of Husserl and Phenomenology [vid. supra, nota al pie 5 cap. 1]. In any case, when we discussed my own work on intentionality he certainly didn't encourage me to follow him in attempting to plumb the depths of the Continental Husserlians" (Dennett, 1994d: 1).

gadores habrán de interpretar tras depurarlo al contrastarlo con registros paralelos –Dennett habla de tres *transcripciones* independientes de los datos–, en un proceso de eliminación de errores y ambigüedades. El más importante de los supuestos será entonces que existe una interpretación intencional del texto así obtenido, es decir, que los sonidos emitidos por los participantes en el experimento (o marcas gráficas o actos conductuales realizados por los mismos) no son meros ruidos carentes de significado, sino que pueden interpretarse como proposiciones que encierran aquello que el sujeto quería decir para significar tal o cual cosa por tales o cuales razones. De este modo cobra el método la objetividad perseguida por Dennett, una clase lingüística de objetividad: objetividad semántica acerca del significado de las proferencias del sujeto. Junto con esta serie de supuestos habremos de conceder que el texto es interpretable como un sincero testimonio de las *opiniones* y *creencias* del sujeto respecto de su experiencia. Esta serie de supuestos podría resumirse señalando que se asume la racionalidad de los sujetos, su capacidad para usar y comprender el lenguaje y, asimismo, que su conducta lingüística está dotada de significado. Cuando el heterofenomenólogo, partiendo de los señalados supuestos y ante el texto ya depurado, se dispone a leerlo e interpretarlo, ¿qué nos dice Dennett que ha de hacer con él? Leerlo como si de ficción se tratara: debe dejar que el propio texto cree el mundo heterofenomenológico del sujeto, poblado por las entidades que los actos de habla de éste permiten al heterofenomenólogo ubicar en él. El texto ha de ser entendido como portador de un cierto tipo de normatividad: hay cosas que cabe preguntarse en vista de lo expuesto en él y cosas que restan indeterminadas más allá del mismo. Lo mismo sucede con las obras de ficción: la pregunta acerca del color de los calcetines que el joven Werther llevaba cuando vio por primera vez a Lotte carece de sentido siempre y cuando Goethe no hiciera constar en su obra nada al respecto. Todo queda determinado dentro del texto e indeterminado más allá del mismo. De este modo, "el mundo heterofenomenológico del sujeto [integrado por los objetos y eventos de los que el sujeto informa] será un postulado teórico estable e intersubjetivamente confirmable, con el mismo estatuto metafísico que, pongamos por caso, el Londres de Sherlok Holmes" (Dennett, 1991: 81 del original, 94 de la traducción).

El objetivo del método heterofenomenológico es el de proporcionar mediante la observación objetiva datos empíricos acerca de la conciencia, datos que, en último término, constituyen el testimonio de lo que el sujeto de la investigación heterofenomenológica *cree* sinceramente que puebla su conciencia –su *mundo heterofenomenológico*–, el testimonio de *lo que al sujeto le parece*, extremo respecto del cual tiene absoluta autoridad –

es decir, autoridad acerca de los objetos y acontecimientos que pueblan su mundo heterofenomenológico, acerca de cómo le parece que son las cosas en ese mundo: el autor del texto dicta lo que hay en el mundo que el texto construye–. Esta autoridad se limita, no obstante, a cómo le parecen al sujeto los moradores de su mundo heterofenomenológico y no alcanza los motivos por los cuales tales moradores le parecen de ese modo: resulta crucial para el método que Dennett ofrece que con él puedan distinguirse descripciones y explicaciones. El sujeto puede describir su mundo heterofenomenológico, y aquí tiene siempre la última palabra, pero no es él quien deba decir por qué suceden como suceden las cosas que en dicho mundo suceden. El método heterofenomenológico requiere distinguir entre las descripciones que el sujeto da de su mundo heterofenomenológico, es decir, del modo en que las cosas le parecen, y las explicaciones o teorías que el propio sujeto pueda aventurarse a formular acerca de las causas y procesos por los cuales las cosas le parecen de ese modo. En este segundo campo el sujeto carece completamente de autoridad. Puede aceptarse pues el enunciado “veo un paraguas” como parte de un texto heterofenomenológico admisible, pero el enunciado “ante el ojo de mi mente aparece un objeto que mi módulo lexico-mnésico reconoce como un paraguas al recibir insumos de mis módulos perceptivos” será interpretado como especulación sin valor descriptivo. El método heterofenomenológico concede plena autoridad al sujeto respecto de cómo le parecen las cosas, no obstante, permanece neutral respeto de lo que pueda suceder realmente dentro de dicho sujeto –incluso neutral frente a la posibilidad de que el sujeto carezca de vida mental de cualquier tipo y sólo produzca sonidos mecánicamente–. El sujeto puede informar acerca de cómo le parecen las cosas –y tendrá de hecho completa autoridad acerca de la adecuación de esta descripción–, pero las teorías que pueda tener acerca de las causas por las cuales le parecen de ese modo o acerca de la naturaleza de la experiencia cosnciente carecen de valor para el heterofenomenólogo –y el sujeto carece totalmente de autoridad acerca de la adecuación de tales teorías.

Habiéndonos aproximado al método neutral para la obtención de datos sobre la experiencia consciente elaborado por Dennett, es momento de introducir esquemáticamente su modelo de la conciencia.

### 3._El modelo de Versiones Múltiples

Una buena manera de aproximarse a la teoría de la conciencia desarrollada por Dennett es partir de la concepción de la misma a la que el de Boston opone la suya. Dennett denomina al objetivo de sus críticas "metáfora del Teatro Cartesiano", la idea de que en la experiencia consciente se presenta todo junto, como proyectado sobre una pantalla situada frente a un hipotético espectador.[7] A pesar del nombre de la referida metáfora no debemos entender que Dennett se propone con su ataque a la misma refutar la filosofía cartesiana de la mente, sino una concepción heredada y endémica cuyos supuestos aparecen velada o explícitamente en las actuales teorías científicas y filosóficas de la conciencia. Que Dennett no ataca directamente la filosofía cartesiana de la mente puede apreciarse ya en el hecho de que uno de los rasgos fundamentales de la metáfora del Teatro Cartesiano es que se trata de una metáfora espacial, cuando Descartes caracterizó la mente incidiendo en el carácter inespacial de la misma. El Teatro Cartesiano de la conciencia se convierte en el blanco de las críticas de Dennett en tanto éste trata de rebatir los modelos espaciales y centrales de la misma: el Teatro Cartesiano sería ese espacio central al que todo llega junto y a la vez para una clara y distinta presentación unitaria ante la audiencia, ante el homúnculo cartesiano. Según Dennett, debemos deshacernos de la idea de este centro: no hay ningún punto en el sistema nervioso central al que lleguen acrisolados todos los resultados de los diferentes cursos de procesamiento de información, no hay escenario para el Teatro Cartesiano. No existe una meta cruzada la cual la información pase de ser inconsciente a ser consciente y presentársenos ordenada y coherentemente.

Frente a esta –supuesta– concepción heredada Dennett defiende un modelo diferente de la conciencia: el modelo de Versiones Múltiples (al que en adelante nos referiremos por sus siglas en inglés: MDM). El punto de partida para la formulación del MDM se encuentra en la idea de que en el cerebro la información es procesada masivamente en paralelo a través de diversas vías en las que la misma es sucesivamente elaborada e interpretada. El procesamiento serial que nuestra ordenada experiencia consciente sugiere es confrontado en el marco del MDM a una vorágine paralela de procesamiento de información llevada a cabo por una enorme cantidad de módulos –sucesivamente más estúpidos, como Dennett subraya en conformidad con su funcionalismo homuncular–, en un proce-

[7] Ya en la década de los cincuenta Ullin T. Place se refería en estos términos a la que denominaba la falacia fenomenológica: "the mistake of supposing that when the subject describes his experience, when he describes how things look, sound, smell, taste, or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen" (Place, 1956: 49).

so en el que la información es reelaborada a la luz de nuevas contingencias, es decir, procesada y revisada con vistas a su integración coherente con la nueva información que constantemente accede al sistema nervioso central. Los poderes que habría que atribuir al centro articulador al que Dennett se refiere como Teatro Cartesiano se distribuyen en el MDM –nuevamente de conformidad con los lineamientos del funcionalismo homuncular– en una serie de agencias sucesivamente menores esparcidas en el tiempo y el espacio neurocognitivo.[8] Así, el trabajo no ha de ser hecho dos veces: una vez llevada a cabo la labor de procesamiento e interpretación que tales agencias realizan no habría necesidad de enviar resultados al Teatro Cartesiano para que allí los estímulos, por ejemplo, vuelvan a ser advertidos, discriminados o disfrutados. Nos hallaríamos ante un flujo masivo y paralelo de procesos simultáneos de fijación y transformación de contenido. De esta manera, dispondríamos en todo momento de gran variedad de borradores del mismo artículo, y la apariencia final, la ilusión según la cual una sola versión del mismo ha existido todo el tiempo se debería a la ocurrencia de sondeos que retrospectivamente favorecen unos borradores en detrimento de otros.[9] Un factor crucial del modelo dennettiano es el tiempo: no existiendo ese lugar al que todo llega junto en que consistiría el Teatro Cartesiano, no puede trazarse una línea en ningún lugar del cerebro en virtud de la cual cupiera distinguir entre procesos conscientes e inconscientes. Intuitivamente, parece que existe un momento en el que se da un tránsito de lo inconsciente a lo consciente. Parece, por ejemplo, que existe un concretísimo momento en que un determinado estímulo pasa del subsuelo del procesamiento inconsciente al teatro de la actividad consciente, pero, careciendo de sentido la idea de que los procesos neuronales relacionados con la percepción de tal estímulo tengan que dirigirse al Teatro Cartesiano para operar allí una suerte de segunda transducción, ¿cómo diferenciar los procesos neuronales inconscientes relacionados con tal estímulo de los que subyacerían ya a la experiencia consciente del mismo?[10] Al no

[8] El homúnculo único que haría las veces de espectador en el Teatro Cartesiano es de este modo partido en mil pedazos que pasarían a ocuparse de tareas menores estableciendo, en el marco de unas determinadas relaciones jerárquicas –en muchos casos de inclusión–, confederaciones que, al ser miradas con lupa, nos llevarían de ligas de homúnculos dedicadas a tareas relativamente complejas a niveles sucesivamente inferiores hasta alcanzar un punto en que un único homúnculo estúpido realiza tareas simplísimas, mecánicas.

[9] Dennett emplea esta terminología tomada del mundo editorial (borrador, artículo) extrayendo jugo retórico del hecho de que en muchas ocasiones, antes de publicar un artículo, diferentes versiones del mismo son enviadas a colegas, los cuales anotan comentarios y realizan correcciones, de forma que en un momento dado circulan diferentes versiones del mismo artículo.

[10] No quisiéramos conducir a errores. Cuando hablamos de procesos neuronales conscientes o inconscientes dentro del marco dennettiano debemos tener presente que para Dennett esos procesos no son propiamente neuronales, sino abstractos, formales: lo que cuanta es la forma de la actividad, no el substrato de la misma. Funcionalista ortodoxo, aprovecha incluso el contexto de una entrevista de carácter divulgativo

existir tal lugar central, como indicábamos, no existe tampoco un momento definido en que un contenido pase de ser inconsciente a ser consciente.

Presentadas las cosas de este modo, según el MDM, por ejemplo, ante un rasgo concreto de una percepción sensorial no ocurre que una representación acabada del mismo sea conducida a un punto central en el que pase ésta a integrarse en una escena consciente coherente, sino que la información correspondiente a tal estímulo viaja por vías paralelas en las que permanece disponible para sucesivas reelaboraciones destinadas a su adaptación a nuevas circunstancias vinculadas con la irrupción de nueva información relacionada. De este modo, diversas versiones de –la información correspondiente a– ese mismo rasgo estarían disponibles en un mismo momento, y la mayoría de ellas estarían destinadas a no jugar ningún papel en la conciencia del sujeto. No puede fijarse ni el lugar ni el momento en el que algo se hace consciente, y el resultado de esos procesos paralelos de elaboración e interpretación de la información, el resultado que llamamos conciencia, es la generación de un flujo narrativo estrechamente vinculado con nuestra dotación cultural y nuestra capacidad lingüística y mnésica, un flujo que puede verse sometido a revisiones del mismo modo que el procesamiento paralelo que hallamos a su base opera constantes revisiones sobre sus contenidos. El MDM establece así que no existe un último eslabón ni un último despacho de revisión en el curso de procesamiento de información que constituye la conciencia.

Puede ilustrarse el significado global del MDM apuntando que diferentes versiones de un mismo contenido circulan tratando de preponderar sin que desde ningún centro articulador se determine la hegemonía de una de ellas, y que logran efectivamente el predominio aquellas que se muestran capaces de monopolizar recursos continuadamente y repercutir en la conducta y la memoria –los contenidos objeto de los referidos sondeos–, en la cual pasarían de este modo a tener la posibilidad de ingresar –pudiendo ser, incluso una vez alcanzado este punto, objeto de nuevas revisiones.

> Los contenidos mentales se hacen conscientes no por ingresar en una determinada cámara especial del cerebro, no por verse transducidos a un medio privilegiado y misterioso, sino por triunfar frente a otros contenidos mentales en el dominio del control de la conducta, y, por ende, de conseguir efectos más duraderos o, como decimos equívocamente, "memorizarlos". Y como somos hablantes, y como hablar con nosotros mismos es una de nuestras actividades más influyentes, una de las formas más efectivas de que un contenido mental se vuelva influyente es que ocupe una posición en la que controle las partes que utiliza el lenguaje (Dennett, 1996a: 155 del original, 183-184 de la traducción).

---

para subrayar que "no son células sino modelos de información" (Dennett, 2004: 174) lo que debe estudiar el investigador de lo mental.

Dennett estaría, en definitiva, planteando que no existe ningún lugar desde el que se lleve a cabo un control centralizado, sino que todo el trabajo sería realizado mediante procesos y subprocesos paralelos y relativamente independientes de procesamiento y reelaboración de la información, es decir, que la información recogida del medio y del propio organismo no es, según el MDM, conducida tras ser recibida y manipulada en las etapas iniciales de procesamiento a un centro rector, cardinal o matricial y tratada allí serial o secuencialmente tras aquella etapa previa de "procesamiento preliminar". Nuestra experiencia consciente sería pues, según esto, el resultado de multitud de procesos interpretativos operados sobre una gran cantidad de versiones de los mismos contenidos, las cuales, como apuntábamos, se disputarían el predominio, una vez alcanzado el cual no pasarían a estar exentas de posibles revisiones ulteriores –incluso después de haber pasado a formar parte del señalado flujo narrativo y haber ingresado en la memoria–. Partiendo de estas premisas, Dennett propone que la conciencia consistiría en un complejo de memes ejecutados en una suerte de máquina virtual tipo Von Neumann implementada en la maquinaria paralela del cerebro –y serían precisamente estos memes y el lenguaje que los sustenta los creadores de dicha máquina *virtual*.[11]

Dennett ha ilustrado recientemente su modelo sustituyendo la denominación MDM por la metáfora "fama en el cerebro" (vid. Dennett, 2005: cap. 6 y 7). La idea esencial consistiría en que, al igual que, según el MDM, no se puede establecer un momento y un lugar en el que un contenido se hace consciente, tampoco puede datarse con precisión el momento exacto en que alguien se hace famoso. La fama en el cerebro es, como el MDM, una metáfora competitiva: los contenidos luchan por la fama, y es evidente que, al igual que sucede con la fama de verdad, no todos pueden ser famosos. Asimismo, al igual que sucede con la fama de verdad, un contenido no puede hacerse famoso unos pocos segundos para luego perderse en el olvido. Eso no sería fama –id est, conciencia– "de verdad", y el medio esencial para conseguirla seguiría siendo la acaparación de recursos lingüísticos y el ingreso en el mundo virtual en que tienen ancoraje los flujos narrativos que nos somos, un flujo que, alzándose sobre el caos paralelo de la actividad cognitiva, simula ordenada serialidad.

[11] El término «meme», como es sabido, refiere a cierta suerte de análogo cultural de los genes. Fue acuñado por Richard Dawkins en la más comentada de sus obras (y quizá también la más comentada dentro de la tradicicón sociobiológica): *The Selfish Gene*, de 1976.

## 4._Los qualia negados por Dennett

*Las ciencias cognitivas deben explicar esas experiencias, no negarlas*. Paul Thagard, 2005

### 4.1._Argumentos de Dennett

#### a) Quining qualia (1988)

Al enfrentarnos a este controvertido artículo de finales de los ochenta nos hallamos ante el primer intento sistemático de remover las más arraigadas intuiciones de la comunidad filosófica y científica interesada en el problema de la conciencia. Muchas de las "bombas de intuición" –*intuition pumps*–[12] que Dennett utiliza aquí por primera vez para desautorizar la noción de *qualia* serán retomadas y reutilizadas en posteriores argumentaciones. Se trata, en definitiva, del lugar desde el cual suele comentarse y analizarse el conato eliminativista de Dennett.

Siendo, como hemos visto, la noción de *qualia* una noción que en absoluto puede considerarse como claramente definida, homogénea o unánimemente aceptada bajo una caracterización estándar, lo primero que cabe preguntarse es qué noción de *qualia* pretende eliminar –*quinear*– Dennett. Su respuesta no se hace esperar: de haber una noción pre-teórica de *qualia* de la cual las definiciones con las que trabajan los teóricos constituyeran simples refinaciones –lo cual no sucede así, dado que se trata de un término técnico; de este modo, cabría sustituir en este punto la locución «noción pre-teórica» por «colección desarticulada de asunciones e intuiciones no justificadas»–, ésa sería la noción de *qualia* a la que estarían dirigidos los dardos de la colección de "bombas de intuición" que el de Boston pone en marcha en este artículo. Prestaremos a continuación atención a los puntos fundamentales de la argumentación que Dennett desarrolla en el mismo para valorar después qué extraemos de las burbujas de intuición que sus dispositivos argumentativos bombean.

La primera de las bombas de intuición (*watching you eat coliflawer*) está destinada a comenzar a cerrar el cerco entorno a las supuestas propiedades especiales de los *qualia*. Imaginemos a una persona a la que el sabor de la coliflor le desagrada en extremo viendo

[12] Dennett utiliza esta noción para referirse a los ejemplos, dispositivos retóricos y experimentos mentales mediante los cuales desarrolla su argumentación.

a otra comer coliflor con deleite. La primera persona podría concluir que la coliflor le sabe a la segunda diferente ya que, después de todo, sabe que en diferentes ocasiones diferentes alimentos le saben de distinto modo. Así, parecería razonable hablar del modo en que, por ejemplo, el zumo de naranja le sabe en determinado momento y preguntarse si es o no diferente del modo en que le sabe en otro. Parece razonable, pero, argumenta Dennett, esto implica la injustificada presunción de que sería posible aislar el modo en que el zumo le sabe en un momento dado y diferenciarlo del resto de concomitancias ocurrentes en ese mismo momento, tales como subsecuentes disposiciones a actuar o juzgar. Desde el punto de vista de Dennett, el error que este tipo de razonamiento implica no es tanto que resulte imposible destilar de esta forma los *qualia* (ese modo central y especial en que las cosas saben, huelen o suenan) y tomarlos en consideración como independientes de tales ocurrencias concomitantes, sino la presunción de que tras semejante destilación habríamos de hallarnos con algo en absoluto, en concreto, con los *qualia*.

La segunda "bomba de intuición" (*the wine-tasting machine*) prepara el terreno para la caracterización que Dennett hace de los *qualia*. En ella hace explícita qué corazonada está tratando de disolver: la de que da igual la capacidad de un sistema para procesar información, y asimismo dan igual las propiedades funcionales y disposicionales que los estados internos de tal sistema puedan tener, él no disfrutaría del modo en que las cosas nos parecen a nosotros, porque nada de eso será especial en el sentido en que los *qualia* son especiales. Lo que hay de especial en los *qualia* Dennett lo rastrea en una sucinta caracterización que, presume –comenta pero no documenta–, es la caracterización tradicional de la noción de *qualia*. Según la misma, los *qualia* serían propiedades inefables, intrínsecas, privadas y directa o inmediatamente aprehensibles de los estados mentales. Veamos brevemente cómo perfila cada uno de estos rasgos, cada una de estas propiedades de segundo orden.

Los *qualia* son inefables porque, a pesar de la elocuencia y la sutileza de los interlocutores, éstos serán siempre incapaces de expresar el particular modo en que ven, oyen, sienten, etc. Además, continúa Dennett, según esta tradición qualófila, una razón que, en parte, explica esta inefabilidad sería que los *qualia* son propiedades intrínsecas, cosa que, dice Dennett, parece implicar que son de algún modo atómicos e inanalizables, homogéneos y simples, y que no habría por tanto forma de describirlos. Así las cosas, los *qualia* aparecerían dentro de esta tradición como privados, y, por tanto, eludirían cualquier intento objetivo de captarlos. Finalmente, siendo como son los *qualia* propiedades de la experiencia, se trataría de propiedades directamente accesibles en la conciencia.

Dennett advierte que esta caracterización sería tan susceptible de ataque como otras, supuestamente alternativas a la misma –por ejemplo, aquellas que aluden a los *qualia* como características fenoménicas de la experiencia, dado que, señala Dennett, la propia noción de fenoménico le resulta nada obvia y tendenciosa y, además, parece remitir o llevarnos de vuelta a la inefabilidad, la privacidad, etc.

Las "bombas de intuición" 3-6 atacan los experimentos mentales basados en espectros invertidos. La versión interpersonal es descartada por cuanto resulta imposible la comparación de la experiencia entre sujetos, y la intrapersonal por cuanto no existiría método introspectivo del que servirse de cara a decidir si la inversión se produjo por alteración del sistema perceptivo o por alteración del sistema de memoria: supuestamente, los efectos experienciales o introspectivos serían equivalentes, cuando sólo en la primera situación los *qualia* se habrían invertido –de forma que sólo análisis externos (neurofisiológicos) podrían determinar dónde se ubicaría la alteración, con lo cual el supuesto del acceso directo a los propios *qualia* se mostraría erróneo, en tanto el sujeto no podría escoger entre las dos posibilidades (alteración perceptiva o mnésica) ateniéndose exclusivamente a la evidencia introspectiva–. Con su ataque a la posibilidad y la plausibilidad de los experimentos mentales basados en espectros invertidos intrapersonales Dennett pretende mostrar que semejantes "bombas de intuición" no bombean las burbujas pertinentes y son insuficientes para probar intuitivamente la existencia de los *qualia* como propiedades directamente accesibles mediante introspección. De este modo, Dennett habría demostrado la falibilidad introspectiva de nuestro conocimiento de nuestros propios *qualia*.

Con la "bomba de intuición" séptima llegamos al punto del artículo más comentado en la bibliografía acerca de los *qualia*: el caso de los catadores de café Chase y Sanborn. Ambos, tras seis años trabajando como catadores para la misma compañía, se informan mutuamente de que el sabor del café de la compañía ha dejado de gustarles. El resto de los catadores están de acuerdo en que el sabor sigue siendo el mismo, de modo que tanto el uno como el otro explican la situación no en referencia a un cambio *objetivo* del –sabor del– café de la compañía, sino en virtud de dos diferentes cambios subjetivos: Chase propone que el café tiene pare él el mismo sabor mientras lo que ha cambiado han sido son sus gustos (se ha convertido en un bebedor de café más sofisticado), y Sanborn piensa, en cambio, que sus gustos permanecen inalterados y que es una variación en su sistema perceptivo la causa de que el café de la compañía haya dejado de gustarle. Expuesto el caso, Dennett nos invita a preguntarnos si cabe la posibilidad de que la explicación que

uno y otro ofrecen sean falsas ambas, o sólo una de ellas. En ambos casos cabría la posibilidad de que los *qualia* gustativos hubieran variado, de que lo que hubiera variado fueran sus actitudes reactivas (*gustos*), o una sutil mezcla de ambas posibilidades. Nuevamente el acceso privilegiado a los *qualia* y la infalibilidad de los juicios que tal acceso propiciaría son contrapuestos a posibles alteraciones mnemónicas. Nuevamente la introspección es incapaz de resolver el rompecabezas: mediante ella resulta imposible para los protagonistas de la bomba de intuición hallar nada decisivo, nada que sirva para optar por una de las tres opciones que se presentan en cada caso (variación de *qualia*, variación de actitudes reactivas o mezcla de ambas). ¿Cómo decidir cuál de las opciones es la correcta? Parece que cabría la posibilidad de averiguar mediante exhaustivos análisis neurofisiológicos si la transformación cae en cualquiera de los dos casos (el de Chase o el de Sanborn) más cerca del procesamiento perceptivo bruto o de los procesos vinculados con los ulteriores juicios reactivos, pero la "bomba de intuición" octava (*the gradual post operative recovery*) está destinada a disolver la intuición de que cabe tal posibilidad. Una intervención quirúrgica invierte las conexiones que partirían de las papilas gustativas de Chase y, en la línea de los clásicos experimentos de espectro invertido, éste se adapta de tal forma a la inversión que su conducta es indiscernible de la de los sujetos normales. Cabría la posibilidad de que su adaptación sucediera con anterioridad o con posterioridad a su experiencia de los correspondientes *qualia* (*pre* o *post-qualia*). En el primer caso los ajustes compensatorios internos destinados a la adaptación a la inversión se producirían antes de que Chase experimentara los *qualia* correspondientes, y en el segundo sucedería lo contrario. Pero, según Dennett, los hechos fisiológicos no podrían servirnos para trazar esta línea divisoria indicándonos en qué fase del proceso surgen los *qualia* como propiedades de esa fase concreta del proceso. Siempre cabrán al menos dos interpretaciones de los datos fisiológicos: supongamos que nuestra teoría neurofisiológica indica que la adaptación se ha producido por un ajuste en el mecanismo de acceso a la memoria que permitiría a Chase comparar sus *qualia* actuales con los anteriores a la operación. En este supuesto caben, como señalábamos, dos posibilidades: que los actuales *qualia* de Chase sean anormales pero él no lo advierta por modificación en su mecanismo de memoria (modificación compensatoria de su memoria acerca del modo en que las cosas solían saberle), y que el paso de la comparación mnemónica ocurriera justo antes de la aparición del *quale* gustativo, y gracias a la revisión el mismo *quale* gustativo se siguiera de la misma estimulación (hemos de entender que de la misma estimulación previa a la operación). En el primer caso los *qualia* formarían parte del *input* a la comparación mnemóni-

ca, y en el segundo serían parte del output de la misma. Se trata de dos hipótesis ciertamente diversas, pero la evidencia neurofisiológica (según Dennett independientemente de hasta qué punto desarrollada) no podrá decirnos en qué lado de la memoria hemos de ubicar los *qualia*. Aquí ya no sólo no puede acceder a sus propios *qualia* certeramente el sujeto que estaría experimentándolos, sino que ambas hipótesis son indiscernibles incluso atendiendo a la evidencia objetiva.

Para finalizar, el ataque a la intrinsecalidad que Dennett propone en este artículo se basa en su famoso experimento mental de los bebedores de cerveza. El gusto por la cerveza es un gusto adquirido: a nadie le gusta su primer trago. Dos bebedores de cerveza proponen dos hipótesis acerca de este proceso: 1) es el sabor de la cerveza lo que cambia, y 2) el sabor permanece constante mientras lo que cambia son las actitudes reactivas de los sujetos, sus gustos. Es decir, ¿cambia el propio sabor con la experiencia, o acaba por gustarnos el sabor de aquel primer trago? Esta "bomba de intuición" estaría, en resumidas cuentas, destinada a desautorizar la propiedad tradicional de la intrinsecalidad, es decir, lo que Dennett está proponiendo con la misma es que ésta –la supuesta intrinsecalidad de los *qualia*– resultaría problemática a la luz de aquélla –la "bomba de intuición" de los bebedores de cerveza–, dado que si aceptamos que el gusto por la cerveza es adquirido, estaremos incluyendo actitudes hacia estímulos y reacciones a los mismos como constituyentes de las cualidades experienciales –de tal modo que éstas se tornarían propiedades relacionales (lo mismo tratan de probar las bombas de intuición 11 y 12, en las que consideramos, por tanto, innecesario detenernos).

b) La conciencia explicada (1991)

Tres años después de "Quining qualia", Dennett, en el que él mismo considera su trabajo fundamental sobre la conciencia, retoma su crítica de la noción de *qualia* proponiendo que ella se nutre de una serie de intuiciones y supuestos no analizados que configuran un cuerpo doctrinal que mantiene a la comunidad filosófica encerrada en el Teatro Cartesiano a pesar de que las paradojas inherentes a dicho círculo de intuiciones hayan sido ya expuestas a la luz pública. En semejante situación, achaca la obstinación por permanecer dentro del Teatro Cartesiano a la ausencia de una concepción alternativa, de la cual dispondríamos, precisamente, desde 1991: el MDM.

Dennett parte en esta ocasión de las llamadas (desde Locke) cualidades secundarias: olores, sabores, colores, sonidos, sensaciones térmicas. La ciencia moderna habría pro-

bado que éstas no se hallan afuera, en el mundo o en las cosas mismas, de forma que, ¿no estarán acaso en nuestras mentes? De hecho el concepto de cualidad secundaria, tal y como Locke lo legara, refiere a la capacidad que los objetos materiales tendrían para provocar determinadas impresiones en nuestras mentes en virtud de sus cualidades primarias (las cuales, a diferencia de las secundarías, se consideran inseparables de la materia, y, asimismo, que nuestra noción de las mismas, nuevamente a diferencia de nuestra noción de las cualidades secundarías, se asemeja realmente a las cualidades habidas en los objetos). Dennett advierte que esta tentadora idea puede inducir fácilmente a error. Por ejemplo, la idea de rojo producida en nuestras mentes por las cualidades primarias de un objeto, ¿es ella misma roja, es decir, está coloreada, o es acerca de «rojo»? La primera opción es tan problemática como la segunda: aquélla requeriría de cierto figmento mental mientras ésta nos impele en una extraña dirección: ¿acerca de qué «rojo»?, ¿el mismo que expulsamos del mundo exterior para convertirlo en una idea acerca de «rojo»? Esta parodia del planteamiento tradicional acerca de las propiedades secundarias le sirve a Dennett para preparar el terreno, conduciéndonos al momento actual en el que los filósofos habrían trocado las cualidades secundarias en *qualia*,[13] en cosas en el observador o propiedades del observador, la existencia de las cuales niega Dennett rotundamente, aunque admite, en un pasaje muy citado, que parece haberlas (lo cual parece contradecir su postura, según la cual no hay ningún fenómeno real de apariencia), y parece que deberían estar dentro, justamente, por haber sido desterradas de ahí afuera por la ciencia moderna. Sin embargo, la propuesta de Dennett es que en lugar de ellas todo lo que hay, es decir, aquello a lo que se reduciría *el modo en que las cosas nos parecen*, serían estados discriminativos de los sistemas nerviosos de los observadores que subyacerían a una amplia gama de disposiciones reactivas. De este modo, cuando hablamos del modo en que las cosas nos parecen, del modo en que las cosas se nos presentan, lo que emitimos no serían sino juicios acerca de tales estados discriminativos, al margen de los cuales resultaría inútil buscar esas adicionales propiedades subjetivas, intrínsecas, privadas e inefables de los estados mentales que los *qualia* –supuestamente– son. Así, los *qualia* se verían reducidos a "la suma total de disposiciones a reaccionar idiosincrásicas inherentes a [nuestros] sistema[s] nervioso[s] como resultado del hecho de que (…) [nos enfrentemos] a un determinado patrón de estímulos" (Dennett, 1991a: 387 del original, 398 de la traducción).

[13] Alfonso García Suárez propone el siguiente matiz: las cualidades secundarías no eran sino cualidades de los objetos de la experiencia mientras los *qualia* son cualidades de la experiencia misma –de la experiencia de los objetos (García Suárez, 1995: 354).

> I claim, then, that sensory qualities are nothing other than the dispositional properties of cerebral states to produce certain further effects in the very observers whose states they are (Dennett, 1994b: 133).

Dennett se centrará en esta ocasión en los colores para desplegar su propuesta eliminativista. Su primer punto consiste en ofrecer una explicación evolucionista de la existencia de los colores según la cual éstos deben su existencia a un proceso evolutivo paralelo de colores (en realidad códigos de colores) y detectores de colores o visión en color, una explicación de la cual cabe colegir que los colores son cualidades cuya existencia se halla inextricablemente vinculada a la de la clase pertinente de observadores de referencia. Partiendo de aquí, Dennett despacha rápidamente los experimentos mentales tipo espectro invertido interpersonal por las mismas razones que los rechazara en "Quining qualia". Las dificultades que implicaría la versión intrapersonal del experimento son ahora presentadas y criticadas con más amplitud. Dennett parte del un supuesto según el cual "lo que necesita el qualófilo es un experimento mental que demuestre que el modo en que las cosas se ven puede ser independiente de todas [aquellas] disposiciones reactivas" (Dennett, 1991a: 391-392 del original, 403 de la traducción).[14] Las versiones del experimento que analiza parten de una hipotética intervención quirúrgica perpetrada por un malvado grupo de cirujanos sobre un inocente sujeto profundamente dormido. Al despertar, el sujeto constata que sus experiencias cromáticas están sistemáticamente invertidas, pero también lo están todas sus disposiciones reactivas –incluida la verbal–. ¿Cómo alcanzar pues la independencia respecto de tales disposiciones que supuestamente necesita el *qualófilo*? La primera opción que analiza Dennett sería un "recableado" mediante una segunda intervención quirúrgica. Con él se reinstaurarían todas las disposiciones reactivas, pero también los *qualia*, a no ser que pudiéramos realizarlo después de que los *qualia* salieran a exhibirse en la conciencia pero antes de que se produjera reacción alguna. La posibilidad es atractiva, pero las consecuencias del MDM la prohíben: según Dennett la posibilidad sólo sería plausible y totalmente legítima sobre las tablas del desmontado Teatro Cartesiano, es decir, si cupiera señalar un lugar en el que ingresan uno tras otro los contenidos de la conciencia, pero según el MDM nada de esto sucede, dado que no

[14] En la entrevista que Gazzaniga hace a Dennett y recoge en las últimas páginas de *Conversations in the Cognitive Neurosciences*, éste pone en boca del qualófilo la misma idea de forma paródica: "«Here am I, looking at the apple, and reflecting on how wonderfully red it appears. Now I subtract my reflections, my dispositions, my changes in mood, my memories, my… and I ask: 'what's left?' and I 'see' that there is still something left over: the very intrinsic redness of it all!» That is not an argument; you couldn't prove anything with such an exercise of the imagination" (Dennett, 1996b: 184).

hay ninguna sucesión discreta de estados mentales conscientes en el escenario del Teatro Cartesiano, sino una marabunta paralela de procesos distribuidos en una red de revisiones constantes que haría imposible trazar con precisión una clara línea divisoria para determinar ese antes y ese después. Esta versión del "recableado quirúrgico" quedaría pues invalidada a la luz de las consecuencias del MDM.

La siguiente posibilidad es la estándar en la bibliografía: el sujeto, simplemente, se adaptaría y todas sus disposiciones reactivas se normalizarían permaneciendo invertidos exclusivamente sus *qualia*. Restauradas sus disposiciones reactivas: ¿qué habría pasado con sus *qualia*? El experimento nos pide que nos los imaginemos aún invertidos, y que así lo hagamos, dice Dennett, proviene de una intuición sustentada en un supuesto erróneo o al menos injustificado: el de que todo reajuste adaptativo de nuestras reacciones caería del lado pos-experiencial, es decir, post-*qualia*. Dennett invita a pensar en un fin de fiesta alternativo, en una forma de concluir el experimento mental que contaría una nueva historia según la cual el sujeto de la inversión va encontrando conforme se adapta que los colores no le resultan tan extraños, y se ve inmerso en situaciones confusas en las que realiza dobles correcciones ("es verde… no, es roj… no, ¡es verde!"). En tal situación tenderíamos a pensar que bien los *qualia* se habrían adaptado o bien se habrían reinvertido, pero nuevamente esta no sería más que una intuición sustentada en la suposición injustificada de que toda adaptación ha de ser pre o post-experiencial,[15] suposición que sólo resulta legítima o para casos extremos o sobre las tablas del Teatro Cartesiano. Vuelve a traer Dennett en este punto a colación el ejemplo de los bebedores de cerveza para ilustrar este extremo, y concluye que para hallar una interpretación que preserve la verdad de alguna de las afirmaciones de los bebedores de cerveza tendríamos destruir los *qualia* para salvarlos, es decir, tendríamos que ir más allá de sus mundos heterofenomenológicos y atender a los eventos acaecidos en sus sistemas nerviosos, y, si encontraramos en ellos algo que sirviera para estos propósitos, afirma Dennett, será sólo porque nos decidamos a reducir el "cómo sabe" a un complejo de disposiciones reactivas –y precisamente por esto habríamos destruido los *qualia* para salvarlos–. De este modo, respecto del espectro invertido, concluye Dennett, la suposición de que existen *qualia* por encima de tales complejos de disposiciones reactivas, *qualia* que podrían permanecer invertidos

[15] Se trata de la misma distinción establecida en "Quining qualia", pero ahora respaldada por el análisis realizado por Dennett (1991a: Segunda parte, cap. V) acerca de la ilegitimidad de la distinción entre revisiones orwelianas y estalinianas.

aun cuando tales complejos permanecieran constantes o invariables, no forma parte más que de un mito: el del Teatro Cartesiano.

c) ¿Son los qualia lo que hace que valga la pena vivir? (2001)

En 2001 Dennett dictó en París una serie de conferencias (las conferencias Jean Nicod) en la que regresara sobre asuntos de filosofía de la mente y ciencias cognitivas. En una de esas conferencias (recogida en el capítulo IV de Dennett, 2005) ofrece nuevos argumentos eliminativistas. Nos referiremos a ellos presentando los dos nuevos puntales intuitivos que Dennett introduce en este nuevo ataque a los *qualia*.

El primero de ellos se basa en el fenómeno de la ceguera al cambio predicho en Dennett (1991a). Se muestran al sujeto dos fotografías casi idénticas, expuestas sucesivamente durante 250 milisegundos cada una e intercalando una pantalla en blanco (la así llamada máscara) entre cada exposición durante un lapso de cerca de 300 milisegundos. Los sujetos tardan normalmente en torno a medio minuto en detectar la diferencia entre las dos fotografías. En el ejemplo que Dennett utiliza en esta ocasión las imágenes muestran una cocina idéntica, salvo por un cambio de color (de blanco a café) en una de las muchas puertas de alacena que aparecen en las imágenes. Dennett pregunta: antes de advertir el cambio, ¿variaron sus *qualia* de color en esa zona? Las posibles respuestas: sí, no, y no sé –porque a) ahora me doy cuanta de que nunca entendí bien qué quería decir con «*qualia*», b) aunque sabía qué quería decir con «*qualia*» en este caso no tengo acceso en primera persona a mis *qualia*, b.1) ¡y la ciencia en tercera persona tampoco puede acceder a ellos!

Cualquiera de las tres respuestas, propone Dennett, resultaría problemática. En el caso de la afirmativa, nos veríamos ante la necesidad de admitir la ocurrencia de cambios grandes y rápidos en nuestros *qualia* que no advertimos, lo que atentaría contra el supuesto del acceso directo, la incorregibilidad y la idea de que somos la única fuente autorizada para hablar de nuestros *qualia*. Según la negativa, apelando al principio de autoridad, como no advertimos cambios en nuestros *qualia*, entonces éstos no habrían cambiado a pesar de los cambios neuronales que pudieran registrarse. Esta respuesta corre el riesgo, apunta Dennett, de trivializar los *qualia* al tratarlos como lógicamente constituidos por lo que juzgamos o advertimos, lo cual haría que se tambaleara el supuesto según el cual ellos son propiedades intrínsecas. Dennett, para acabar, apunta escuetamente que la tercera respuesta, por su parte, dejaría a los *qualia* en la paradójica situación de ser

inaccesible tanto desde la primera como desde la tercera persona. Según Dennett, este experimento informal probaría que los filósofos no saben muy bien a qué se refieren con el término técnico «*qualia*».

El segundo de los nuevos puntales intuitivos dennettianos es el experimento mental del señor Clapgrass, un sujeto habitual de experimentos neuropsicológicos que un día, al despertarse, comprueba horrorizado que el mundo le resulta extraño. Los científicos con los que lleva meses trabajando le preguntan qué ve al abrir los ojos, a lo que responde nombrando apropiadamente los colores de los objetos (cielo azul, hierba verde, etc.). Realiza las pruebas estándar para detectar el daltonismo (la prueba Ishihara) u otros desórdenes en la percepción del color, pero el señor Clapgras identifica, discrimina y denomina correctamente los colores, con lo cual todo el equipo científico está de acuerdo en que cualquiera que sea su problema no implica a la visión en color, todo el equipo excepto el doctor Cromófilo, que durante meses ha recopilado datos acerca de reacciones fisiológicas sutiles relacionadas con la respuesta emocional a los colores del señor Clapgras e insiste en realizar algunas pruebas más. Al llevarlas a cabo descubre que las reacciones fisiológicas relacionadas con respuestas emocionales al color de Clapgras se hallan sistemáticamente invertidas. ¡Ha sufrido una inversión de las reacciones emocionales y atencionales al color, una inversión de los efectos subjetivos que los estímulos cromáticos solían tener en él!, una inversión de sus *qualia* cromáticos que, sin embargo, ha dejado intactas sus capacidades de reconocimiento, identificación, discriminación y denominación. ¿Qué ha pasado con sus *qualia*? Dennett plantea que su experimento mental, frente a los habituales en la bibliografía, contiene la innovación de que en el caso del señor Clapgras, sus *qualia* podrían haberse invertido sin que él se enterase de nada. Según Dennett, el principal problema para los filósofos estaría en que han venido suponiendo que el tipo de cosas que tanto el doctor Cromófilo como el resto del equipo ha detectado en el extraño caso del señor Clapgras no tendrían nada que ver con los *qualia*, mientras que lo que tanto a él como al resto del equipo les parece que ha pasado con lo que habitualmente los filósofos entienden por «*qualia*» ha permanecido inalterado. Por lo tanto, la conclusión, desde el punto de vista de la noción de *qualia* habitual en la bibliografía, habría de ser que los *qualia* del señor Clapgras no han variado. Su *visión de los colores* es perfecta (tanto él como el equipo científico están convencidos de ello), pero para asegurarse de esto el desconcertado Clapgras ha tenido que apoyarse en baterías de pruebas. El problema, en cualquier caso, estaría nuevamente en la memoria y la imaginación, dado que los filósofos han venido suponiendo que la imaginación de un color y la

comparación en la memoria o el recuerdo del mismo son procesos claros y aproblemáticos, tendiendo a pensar así que uno puede estar seguro de que lo que hace al imaginar el color amarillo es lo mismo que ha hecho siempre al imaginar ese color.

Dennett aclara que no se ha referido al presentar el extraño caso del señor Clapgras a descripciones del estado subjetivo del señor Clapgras respecto de sus experiencias cromáticas: no ha dicho cosas tales como que cuando mira el cielo experimenta *azul subjetivo intrínseco*, precisamente porque lo que pone en duda es que esos términos se refieran a verdaderas propiedades de la experiencia, dado que restan, en vista de lo dicho, más allá del alcance de los datos objetivos de tercera persona, pero también más allá del alcance introspectivo del propio señor Clapgras, con lo cual Dennett no encuentra sitio en el estudio científico de la conciencia para estas propiedades intrínsecas de los estados mentales conscientes.

4.2._Problemas que suscitan los argumentos de Dennett

a) En “Quining qualia” Dennett parte del supuesto según el cual nuestra noción intuitiva de *qualia* es un vacuo galimatías, y en tanto entiende que la concepción tradicional que en filosofía de la mente recogería esa noción intuitiva se encuentra en consonancia con la misma, trata de matar dos pájaros de un tiro socavando a la vez ese concepto intuitivo y esa concepción tradicional mediante la siguiente estrategia: si las cuatro propiedades tradicionales no se sustentan, llegando a resultar mutuamente incompatibles, entonces, nuestra concepción intuitiva de los *qualia* ha de ser incoherente y, por tanto, el término «*qualia*» carecería de extensión. Nosotros hemos criticado el supuesto según el cual la propiedad de la intrinsecalidad forma parte de una noción de *qualia* aceptable desde un punto de vista naturalista, y creemos que puede igualmente defenderse que, desde el punto de vista de la actitud natural, esto es, desde un punto de vista intuitivo, nadie estaría tentado de integrar tal propiedad en su noción cotidiana de olor, sabor o estado anímico, porque puede que la expresión “aislar el modo en que las cosas me parecen” lleve a muchos a pensar en una posibilidad real, pero nuestra psicología popular no deja de incluir ideas tales como que mi desagradable sensación dolorosa, supuestamente intrínseca según la tradición qualófila, está de hecho *relacionada* con el golpe que he recibido, con mis creencias y reacciones relacionadas con sensaciones dolorosas, etc. Cabe, pues, dudar que la intrinsecalidad forme parte a la vez de la concepción tradicional y de la cotidiana. Parece correcto, además, aceptar que la noción intuitiva difiere de la

tradicional en que la primera podría caracterizarse con expresiones tales como "el modo en que las cosas nos parecen", mientras que la tradicional hace uso de conceptos técnicos de dudosa traducción al terreno cotidiano e intuitivo. Las cuatro propiedades tradicionales a las que Dennett apela y que con su argumentación ataca no parecen buenos candidatos al título de compendio de nuestras intuiciones preteóricas acerca de nuestra experiencia consciente. De este modo, el desmantelamiento de esas cuatro propiedades no sería suficiente para probar la incoherencia de nuestra concepción intuitiva de los *qualia*. ¿Es pues, meramente, la creencia sustentada por los sujetos de que tienen experiencias en las cuales las cosas les parecen de determinados modos lo que debe una teoría científica de la conciencia explicar o, más bien, siguiendo la invitación contenida en la cita de Thagard que encabeza esta sección, esas experiencias mismas?

Los experimentos mentales que Dennett propone en el artículo que venimos comentando están destinados a mostrar la invalidez de alguna de las cuatro propiedades tradicionales. Así, por ejemplo, ataca la propiedad de la aprehensibilidad directa o inmediata mediante sus ajustes en los clásicos experimentos de espectro invertido y su "bomba de intuición" de los catadores de café: nada pueden hallar en su experiencia directa que les sirva para distinguir una inversión de espectro de una inversión mnemónica, del mismo modo que nada pueden hallar en su experiencia directa que les sirva para establecer si lo que han variado han sido sus *qualia* o sus disposiciones a reaccionar ante los mismos. En cualquier caso, los individuos no pueden decidir entre las hipótesis alternativas que pudieran explicar qué ha pasado con sus *qualia* (¿han permanecido idénticos, ha variado su memoria?), pero de hecho los sujetos experimentan variaciones en su experiencia actual: hay para ellos algo desconcertante que tiene, en efecto, que ver con el modo en que las cosas les parecen, y este desconcierto, independientemente de que alcancen o no a estipular si ha tenido lugar una inversión actual o un fallo en la memoria, arraiga en el hecho, un tanto obvio, de que tienen experiencias. Dennett trata de erosionar el supuesto del acceso directo a los *qualia* invitándonos a contemplar casos en los que el acceso directo a la experiencia actual se mezcla con un hipotético acceso confuso a la memoria de *qualia* relacionados con la experiencia actual. La objeción en este punto podría resumirse haciendo notar, por una parte, que el acceso directo, la inmediata presencia no inferencial de la experiencia, no implica acceso hialino y milimétrico a un mundo dado y acabado de infladas e injustificadas certezas *naïve*, por otra parte, que acceso directo y comparación en la memoria no son nociones intercambiables, y, para terminar, que su intrincada interrelación no dice nada acerca de la primera: la falibilidad de mi memoria puede efectiva-

mente tener que ver con el hecho de que confunda sucesos actuales de una clase con sucesos pasados de otra, pero la falibilidad de mi acceso directo a las cosas tal y como ahora las experimento difícilmente puede hacerse depender de mi memoria: ella puede llevarme a decir cosas extrañas, como que el azul que experimento hoy es el mismo que el que experimenté ayer, y podría estar muy equivocado en esto, pero dada la propiedad de la inefabilidad, no sería por otra parte muy claro qué estaría tratando de decir con esto. ¿Puedo creer que no me duelen las muelas cuando en verdad me duelen? ¿Es el estado de creencia constitutivo del dolor? ¿Le dolerían las muelas a un niño salvaje incapaz de hacer fluir memes en su máquina Von Neumann virtual? ¿Es la creencia del dolor el propio dolor? El modo en que Dennett trata de zafarse de esta objeción consiste en responder afirmativamente. Valore el lector si el precio que de este modo paga es o no excesivo.

Cuando Dennett intenta conducir nuestra intuición hacia el rechazo de la aprehensibilidad inmediata, lo que logra es ponernos al corriente de nuestra dudosa capacidad para comparar en la memoria experiencias actuales con experiencias pasadas. Dennett contempla la posibilidad de una objeción de este tipo, es decir, de una objeción según la cual aunque introspectivamente no pueda elegir entre las posibles hipótesis, esto nada diría contra el hecho de que el sujeto detecta cambios y tiene una peculiar experiencia a la cual sí puede acceder directamente –independientemente de que pueda compararla certeramente en su memoria o no–. Contempla la posibilidad, como indicábamos, pero sólo apunta que dicha tentativa de escapatoria no garantiza el éxito del *qualófilo*, porque de hecho nada se sigue de tal objeción. Puede que nada se siga en el sentido de que no ayuda a escoger entre las posibles hipótesis, pero sí hace posible la puesta en tela de juicio la concepción de la aprehensibilidad directa que Dennett estaría manejando en su argumentación, es decir, que cabe desde esta objeción señalar que Dennett no utiliza claramente dicha noción, sino que de hecho mezcla en ella diferentes extremos, llegando a contaminar una noción preteórica e intuitiva de aprehensibilidad directa, que tendría que ver con la experiencia actual, directa, insertando en la misma aspectos no tan directos –tales como la imposibilidad de una certera comparación en la memoria–. Por nuestra parte, al presentar a los *qualia* como directamente accesibles no dijimos nada de infalibilidad a la hora de comparar un *quale* presente con otro pasado, sino que simplemente indicamos que, con arreglo a tal propiedad de acceso o aprehensibilidad directa, los consideraríamos propiedades primarias por cuanto no necesitan para su actualización de derivación mediante razonamiento o inferencia y en tanto no se coligen de otros contenidos sino que se nos dan de forma directa. No es ésta la noción de aprehensibilidad directa que la argu-

mentación de Dennett logra arrinconar y, consideramos, se trata de una concepción de tal propiedad más cercana a ese núcleo de intuiciones desarticuladas e injustificadas que conformarían la noción intuitiva de *qualia* que Dennett se propone impugnar, dado que presenta a los *qualia* como lo que habitualmente se traduce por "el modo en que las cosas me parecen", es decir, el modo directo, no inferencial, en que experimentamos nuestras experiencias.

El ataque al resto de las cuatro propiedades ocupa un lugar secundario en este artículo de Dennett (por ejemplo contra la inefabilidad dice poco más allá de que con la experiencia alcanzamos de hecho la posibilidad de expresar con mayor precisión nuestras experiencias), excepto, tal vez, por el ataque a la intrinsecalidad ensayado mediante el experimento mental de los bebedores de cerveza. Dado que, como Dennett, hemos rechazado esa supuesta propiedad de segundo orden, no entraremos a comentar aquí este experimento mental.

b) Por otra parte, y retomando lo apuntado por Dennett (1991a) acerca de la necesidad en la que se vería el defensor de la noción de *qualia* de diseñar un experimento mental en el que se hiciera obvio que los *qualia* pueden ser de hecho independientes de cualesquiera disposiciones reactivas, cabe señalar que, partiendo del rechazo de la intrinsecalidad como propiedad de los *qualia*, esa necesidad no sería tal. Que los *qualia* hayan de ser netamente independientes de cualesquiera estados del sistema nervioso y de cualesquiera disposiciones a reaccionar es algo que debiera sonar extraño a todo filósofo de la mente que pretenda integrar los *qualia* en un marco teórico naturalista, pues, ¿qué tipo de propiedades podrían permanecer aisladas de todo cuanto causa o está relacionado con el resto de hechos de algún modo asociados con nuestra vida mental? Dennett (1991a) trata de reducir los *qualia* a disposiciones reactivas, mas, una vez sentado que no hablamos de entidades al hablar de ellos, sino de propiedades, y una vez sentado que no hablamos de propiedades intrínsecas al referirnos a ellos, *¿qué nos impide pensar que precisamente esas disposiciones reactivas de un sistema nervioso inmerso en el flujo de su interacción con su cuerpo y su entorno estén dotadas de esas complejas propiedades que han venido designándose con la voz «qualia»?* El intento eliminativista de Dennett (1991a) tiene precisamente a la intrinsecalidad por objetivo principal, pero al atacar esta supuesta propiedad da por supuesto que una vez rechazada la intrinsecalidad y contemplados los estados mentales que supuestamente portarían *qualia* como meros conglomerados reactivos del sistema nervioso, la experiencia consciente no tendría ya ningún lugar en el que es-

conderse. Su argumentación, no obstante, no prohíbe que tales conglomerados porten propiedades fenoménicas. La argumentación de Dennett, a pesar de que se presente como tentadora para un estudio simplificado y naturalista de la conciencia, parece descartar más datos que prejuicios provenientes de una concepción obsoleta de la mente y la conciencia, una concepción que, tal y como Dennett propone, tendería a inflar injustificadamente ambas nociones animada por nostálgicas doctrinas acerca de la magia de la conciencia y –parafraseando a Scheller– el puesto del hombre en el cosmos.

¿Necesita verdaderamente el *qualófilo* probar la independencia de los *qualia* atacada por el núcleo de la argumentación de Dennett (1991a) una vez desechada la propiedad de la intrinsecalidad?

c) Por lo que toca a los dos nuevos puntales a los que nos referimos en último lugar en el apartado anterior, apuntemos brevemente que en el caso del experimento informal basado en la ceguera al cambio, la respuesta negativa que Dennett despacha en unas pocas líneas ni resulta implausible ni trivializa los *qualia* del modo en que Dennett pretende. Si decimos que los *qualia* no variaron *porque* no los advertimos, no estaríamos incluyendo lógicamente en la noción de *qualia* nuestro juicio sobre ellos como constituyente de los mismos, porque de hecho podríamos responder igualmente de forma negativa y decir que no se trata de que no variaran *porque* no los advertimos, sino que no los advertimos *porque* no variaron. De todos modos, nos sobra jugar a esta clase malabarismo léxico tratando de impugnar la relacionalidad de los *qualia*, pues precisamente no tenemos ningún problema en concebirlos como propiedades *efecto de relaciones* y *capaces de causar efectos y entrar en relaciones*. No obstante, nuestra defensa de los *qualia* como propiedades primarias nos fuerza a contemplar que la única respuesta válida es la negativa, la cual no resulta en absoluto problemática: muchos conglomerados reactivos de nuestros sistemas nerviosos nos pasan inadvertidos y, de acuerdo con hipótesis como la del núcleo dinámico, cabría afirmar que hasta que no entran en determinados tipos de relación con los procesos neuronales pertinentes no pueden hacerse conscientes sus contenidos. Una vez integrados en tales flujos de actividad neuronal relacionada con los estímulos que incidieran sobre esa parte de la retina podrían pasar tanto a tener efectos sensibles como a formar parte de nuestros juicios (sobre los cuales la evidencia empírica disponible obliga a convenir en que pueden influir asimismo en ausencia de efectos sensibles): hasta ese momento sólo habría actividad neuronal inconsciente –aunque esto parezca contradecir las implicaciones del MDM, podría integrarse en el mismo (siempre y cuando estuviéra-

mos dispuestos a aceptar la parte más problemática del modelo de la conciencia que propone Dennett: el papel del lenguaje, los memes y la máquina virtual tipo Von Neumann) mediante el siguiente movimiento: ese "hasta ese momento" consistiría en un "hasta que el sondeo correspondiente integrara el procesamiento inconsciente en un fragmento narrativo triunfador o preponderante".

En cuanto al extraño caso del señor Clapgras, incidamos en que 1) también nosotros rechazamos la propiedad de segundo orden de la intrinsecalidad, de modo que poco tendremos que comentar al respecto, y 2) ya hemos criticado la eficacia de la comparación de *qualia* en la memoria con *qualia* actuales para socavar el supuesto de la aprehensibilidad directa, aunque, de todos modos, cabe conceder que en esta ocasión Dennett trata más bien de atacar el supuesto del acceso privilegiado (pues parece que el experimento prueba que Clapgras necesita preguntar y apoyarse en pruebas para saber qué ha pasado con sus *qualia*). Pero, ¿qué ha pasado con sus *qualia*? Esto nos da pie para romper una nueva y nuevamente tangencial lanza en favor de una visión holista de la experiencia consciente aparejada al rechazo de una concepción de los *qualia* que los presenta como caracterizados por la simplicidad y la atomicidad. Un *quale*, si cabe hablar así, por puntual que sea, nunca aparece aislado y simple, como de una sola pieza, sino inextricablemente entremezclado con diferentes aspectos de la experiencia, y ello incluso a pesar de que la modulación atencional alcance a elevarlo, si es que cabe utilizar en este punto este singular, muy por encima del resto de la corriente de la conciencia: no vemos color sin forma, ni creemos en el Ratoncito Pérez en abstracto, sin determinada tonalidad afectiva, rango hedónico o estado de alerta, ni sentimos pena sin que ésta esté integrada, en definitiva, en una vida mental momentánea, puntual y transitoria pero siempre compleja e integrada. Esta integración de lo complejo está constituida por partes, por segmentos de contornos, las más de las veces, extremadamente difusos, y cada una de estas partes puede variar, e incluso desaparecer. Desde este punto de vista, es fácil responder a la pregunta que formula Dennett: los *qualia* cromáticos de Clapgras habrían variado parcialmente, habrían variado por lo que a su aspecto emocional tocaría –sin necesidad de recurrir a los malabarismos terminológicos impuestos por un intento eliminativista que se presenta como el de desterrar un término superfluo pero que deriva finalmente en el de movilizar el asenso respecto del carácter ilusorio de la propia experiencia consciente, podría expresarse esta idea de forma mucho más sencilla: la experiencia consciente del señor Clapgras, efectivamente, ha variado, concretamente por lo toca que a determinado sector de su economía afectiva relacionada con la percepción visual.

## 5._Comentario crítico sobre el método neutral de Dennett

La primera perplejidad que el método heterofenomenológico suscita surge al comprobar que el mismo nos propone tratar como creencias las experiencias de las que informan los sujetos, una estrategia que ha sido criticada, entre otros, por Dreyfus & Kelly (2007), partiendo en su caso del hecho de que el método heterofenomenológico atribuye creencias a los sujetos que meramente informan acerca de sus experiencias. Según Dreyfus y Kelly, el heterofenomenólogo estaría insertando injustificadamente con su interpretación de los informes de los sujetos contenido intencional ausente en la mente de éstos. Esta crítica podría resumirse –excesivamente– indicando que creer no es idéntico a experimentar: una creencia sobre una experiencia consciente dada no equivale a la misma. En nuestra opinión, más allá de la contaminación intencional y el vaciamiento experiencial denunciados por Dreyfus y Kelly, las complicaciones que para el método dennettiano se desprenden de este tratamiento de la experiencia en términos de creencia o juicio tendrían que ver, meramente, con la completa heterogeneidad de tales nociones. Así, por ejemplo, tener una creencia implica conceptualmente otras muchas con diversos contenidos, es decir, tener una creencia determinada conlleva que el sujeto de la misma habría de comprometerse, por el solo hecho de tenerla, con la validez de muchas otras, pero no puede decirse que lo mismo suceda con las experiencias. La experiencia consciente no porta necesariamente valores de verdad y no obliga por tanto a compromiso alguno con una validez que resulta así difícil concebir como en algún sentido imputable a la misma. En otras palabras, cortar la experiencia consciente con el patrón de las actitudes proposicionales tiene sus inconvenientes, y ello porque, sencillamente, no contamos con muchos motivos más allá de determinadas inercias dogmáticas para identificar un dolor de estómago con una proposición o una actitud hacia una proposición. Veamos un solo ejemplo. Cuando afirmo: "opino que la regulación política de los mercados financieros se hace necesaria para afrontar los desafíos de la globalización", esto implica que opino que la política es un extremo acerca del cual cabe opinar, que creo que existe algo denotado por locuciones como «mercados financieros» o voces como «globalización», etc. En cambio, cuando digo que me duele la cabeza no sucede nada parecido pues, aunque una gran cantidad de implicaciones del tipo de las anteriores pudiera igualmente considerarse que deriva de mi aserción o es implicada por la misma (que creo que existen cabezas, y dolores, que las cabezas pueden doler, etc.), al tiempo, y a diferencia del caso anterior, tam-

bién una gran cantidad de consecuencias difícilmente explicitables de forma lingüística podría considerarse como implicada por mi aserto, y de entre esas consecuencias, cabría diferenciar las relacionadas con eso que es para mí como sentir ese dolor de las que pudiera extraer un organismo que comprendiera mi lenguaje y compartiera mis modalidades perceptivas, propioceptivas, nociceptivas, etc., y, por último, de las que pudiera extraer un organismo que comprendiera mi lenguaje pero no compartiera mis modalidades perceptivas, propioceptivas, nociceptivas, etc. Una creencia, un juicio, es, en definitiva, la clase de cosa que evolucionamos para compartir, una actividad inevitablemente dependiente de nuestra capacidad para hablar y utilizar conceptos. Una vez enunciado el juicio, la pérdida de contenido en su viaje hacia el receptor o receptores del mismo puede deberse a la ambigüedad de las expresiones utilizadas, fallos en el medio, etc., pero una vez purgadas esas dificultades o deficiencias no habría inexorables motivos de principio para que incluso un extraterrestre capaz de entender el idioma en el cual fuera formulado el juicio en cuestión sacara las conclusiones apropiadas de lo expuesto en el mismo –es decir, para que nada de lo en él expresado, nada de lo por él prohibido o permitido, concedido o rechazado, afirmado o negado restara desatendido o de algún modo inaccesible–. Una experiencia, por su parte, es la clase de cosa que evolucionamos para advertir o sentir y, congruentemente, no ingresa en la esfera de lo comunicable con tan buenas cartas como las del juicio. Muchos de los moradores de cualquier mundo heterofenomenológico son designados mediante etiquetas que requieren para ser comprendidas de mundos fenomenológicos *en* los intérpretes. Cuando el sujeto dice experimentar un fuerte dolor de estómago está haciendo uso de conceptos cuya comprensión requiere algo más allá de la adquisición proposicional de saber-qué, requiere, en concreto, de una experiencia de la clase pertinente, una suerte de saber-cómo (se oye, se ve, se siente) que el método heterofenomenológico obvia al tiempo que da por supuesta, pues sin ese tipo de saber-cómo la interpretación de la práctica totalidad de cualquier texto heterofenomenológico se asentaría en *flatus vocis*.

Por otra parte, incidir en la diferencia entre experiencia y juicio o creencia no equivale, a nuestro juicio, a inflar la conciencia o a argumentar en la línea de los nostálgicos de la magia de la conciencia. La carga de la prueba está en el campo de Dennett: no vale afirmar que quien pretenda que haya en nuestra economía mental algo aparte de juicios o creencias es porque parte de que la conciencia debe ser algo mágico; no vale afirmar, hay que probar que no hay nada más allá de juicios o creencias, lo cual parece tan difícil co-

mo probar que efectivamente lo hay, aunque sólo si entendemos «probar» como enfrentar a un ejercicio erístico tipo Aquiles y la tortuga otro análogo.

Así las cosas, ¿de qué modo puede el método heterofenomenológico servirnos para partir de lo que el sujeto dice hacia lo que el sujeto experimenta? Ofrecer respuesta a esta pregunta resulta difícil desde la perspectiva ofrecida por la propuesta metodológica dennettiana: la propia formulación de la pregunta choca en este marco metodológico con el hecho de que el acto interpretativo del experimentador consistirá en cualquier caso en la formulación de una hipótesis teórica. El texto heterofenomenológico genera un constructo teórico, el mundo heterofenomenológico del sujeto, y la posibilidad de transitar desde este mundo teórico hacia la propia experiencia del sujeto queda desde el principio descartada por cuanto se la hace coincidir con el mismo.

Dennett admite que parece haber *qualia* (lo cual, ciertamente, parece contradecir su idea de que no hay ningún fenómeno real de apariencia). Lo parece, pero nada más, pues según él no los hay: no hay nada más que juicios, creencias erróneas acerca de un ilusorio mundo de experiencias conscientes, de modo que una metodología adecuada para el estudio de la conciencia habrá de tratar esas supuestas experiencias en términos de juicios, como si sólo hubiera complejos disposicionales que en último término contaran como juicios. Su método en tercera persona parece destinado a tomarse en serio los datos de primera persona ofrecidos por los sujetos, pero sólo lo parece, nada más, porque en verdad, al tratar los informes sobre experiencias como juicios, lo que Dennett está proponiendo es dar una vuelta de tuerca que convertiría los datos de primera persona en datos de tercera persona. Dichos informes harían referencia en su propuesta no a mundos subjetivos realmente existentes, sino a creencias acerca de supuestos mundos subjetivos, a ficciones teóricas. Dennett ataca así la idea de la existencia de experiencias que no creamos tener. Esta identificación de experiencia y juicio, anotemos tangencialmente para acabar, encaja mal en contextos intensionales, y ello, nuevamente, a causa de la heterogeneidad de la aplicabilidad de la noción de validez a las de experiencia y juicio. Asumida la identidad entre juicios o creencias y experiencias, situaciones como la siguiente se vuelven problemáticas: afirmar que X experimenta pavor es necesariamente igual de cierto que afirmar que X experimenta temor, pero afirmar que X cree que experimenta pavor no es necesariamente igual de cierto que afirmar que X cree que experimenta temor – pues sería una falsa atribución en el caso de que X desconociera el significado del término «temor».

Cabe admitir, con Dennett, que mi experiencia consciente no es un pleno, que no hay realmente en ella ni la mitad de lo que me parece, pero ello no es suficiente para hacernos comulgar con un método como el heterofenomenológico. En el fondo, el verdadero problema de dicho método es que se presenta como un método para el estudio de la conciencia pero termina por ser una estrategia retórica al servicio de una metafísica de la conciencia según la cual ésta es una ilusión "linguógena". Precisamente el modo en que esta metafísica hace depender la naturaleza de la experiencia de la del lenguaje[16] –un modo más acusado que el de los teóricos representacionalistas de orden superior, pues Dennett entiende que no es necesario introducir entre la experiencia consciente y la disposición o capacidad de transmitirla lingüísticamente ningún intermediario, porque de hecho la idea de un pensamiento formulado con independencia del lenguaje le parece incoherente– ha sido a menudo criticado como chauvinista en tanto negaría la conciencia a los animales no humanos –vid., v. g., Flanagan (1992); Block (1993); Churchland (1995); Churchland (1998); Clark (2002)–. En cualquier caso, y en la más caritativa interpretación, la heterofenomenología sería una especie de recordatorio de aquello que no es necesario recordar a los investigadores en psicología de la atención o psicología de la percepción: que los sujetos son ingenuos y tienden a sobrestimar sus capacidades introspectivas y la acabada plenitud del mundo al que dan acceso.

[16] "[Consciousness] arises when there is work for it to do, and the preeminent work of consciousness is dependent on sophisticated language-using activities" (Dennett, 1983: 384).

# CAPÍTULO 12

## LA NEUROFENOMENOLOGÍA

«Neurofenomenología» fue originalmente una signatura acuñada de forma nada inocente: con ella pretendía Varela (1996) distanciarse de la neurofilosofía de los Churchland. [1] El marchamo "filosofía" estampado en las aduanas neurofilosóficas de Norteamérica, entendía Varela, resulta un tanto unilateral, pues sólo habla de la filosofía de la mente angloamericana, de los herederos del análisis filosófico, mientras él quería recuperar para las ciencia cognitivas y la reflexión filosófica contemporánea acerca de la mente tanto esas tradiciones que en los manuales de historia de la filosofía venían denominándose "continentales" como aquellas otras que rara vez aparecen en los mismos: las "orientales" (particularmente budistas). Frente al rechazo neurofilosófico de nuestras intuiciones de sentido común acerca de lo mental, los neurofenomenólogos vendrían a adscribirse a aquella famosa fórmula einsteiniana según la cual toda ciencia no es nada más que un refinamiento del sentido común y las operaciones cotidianas de la mente (Einstein, 1936: 290τ). En los siguientes apartados veremos de qué modo plantean los neurofenomenólogos la necesidad y la posibilidad de dicho refinamiento en el caso del desarrollo de una ciencia de la conciencia.

[1] El término «neurofenomenología» sería, no obstante, usado por primera vez en Laughlin, McManus & d'Aquili (1990), un texto en el que se trazan también puentes interdisciplinarios, pero que tienen aquí más que ver con una aproximación multidisciplinar a la relación entre cultura simbólica, mente y cerebro. La acuñación del término obedecería en el caso de este texto al intento de vincular tradición fenomenológica husserliana, algunas prácticas budistas (al igual que en Varela, Thompson & Rosch, 1991), diversas disciplinas antropológicas y neurociencias.

El impulso inicial para el proyecto neurofenomenológico de elaboración de una metodología capaz de tender puentes entre las tradiciones de introspección disciplinada (europeas y orientales) y las modernas ciencias cognitivas tiene un nombre propio: el de Francisco Javier Varela García (Santiago de Chile, 1946 – París, 2001). A él pronto vendrían a sumarse, entre otros, Evan Thompson, Walter Freeman, Shaun Gallagher o Antoine Lutz. Dicho proyecto surge, en cualquier caso, como un segmento del enfoque enactivo en ciencias cognitivas propugnado a comienzos de la década de los noventa en un libro que ha acabado por convertirse en un documento de referencia dentro de la joven tradición neurofenomenológica: *The Embodied Mind*, firmado por Varela (a la sazón y hasta su temprana muerte en 2001 director de investigación del grupo de Neurodinámica en el laboratorio de Neurociencias Cognitivas e Imágenes Cerebrales del CNRS, enclavado en el famoso Hospital de la Pitié-Salpêtrière), Evan Thompson (que mientras redactaba su tesis doctoral en filosofía para la Universidad de Toronto trabajó con Varela en el Centre de Recherche en Epistemologie Applique de la Ecole Polytechnique de Paris) y la entonces ya veterana psicóloga neoyorkina Eleanor Rosch. En este texto el enfoque enactivo aparece como una propuesta de viraje paradigmático en ciencias cognitivas desde la abstracción centralista computacional propia de los paradigmas simbólico y conexionista hacia una consideración de la cognición como esencialmente vinculada con el cuerpo y la acción de los organismos, un viraje, asimismo, hacia una consideración de la mente como mente vivida, experimentada.[2] Es justamente aquí donde el enfoque metodológico neurofenomenológico entraría en juego: la referencia a tradiciones budistas y fenomenológicas en *The Embodied Mind* vendría a prefigurar el posterior desarrollo del programa neurofenomenológico al incidir en la relevancia para el progreso de las ciencias cognitivas de la obtención de informes fenomenológicos rigurosos, que habrían de proceder de sujetos entrenados en diversas tradiciones intros-

[2] Un importante aspecto del señalado viraje es la invitación al abandono de la noción de representación, clave para los dos paradigmas anteriores en ciencias cognitivas. La cognición no es en el enactivismo concebida como algo para cuya explicación se haga necesario el recurso a dicha noción y, de hecho, la misma es mirada con recelo y desconfianza por los partidarios de dicho enfoque por cuanto oculta tres supuestos a los que aconsejan aproximarse con toda cautela: 1) que habitamos un mundo con propiedades particulares, 2) que "captamos" o "recobramos" estas propiedades representándolas internamente y 3) que un "nosotros" subjetivo separado es quien hace estas cosas. Lo que, en resumidas cuentas, el enactivismo vendría a defender frente a los anteriores paradigmas en ciencias cognitivas sería "que la cognición no es la representación de un mundo pre-dado por una mente pre-dada sino más bien la puesta en obra de un mundo y una mente a partir de una historia de la variedad de acciones que un ser realiza en el mundo" (Varela, Thompson & Rosch, 1991: 9 del original, 33-34 de la traducción). Además, desde este prisma, se "sostiene que la cognición depende de los tipos de experiencia que provienen del hecho de tener un cuerpo con varias habilidades sensoriomotrices; y que estas habilidades individuales se alojan a su vez en un contexto biológico y cultural más amplio" (Ojeda, 2001: 291).

pectivas concebidas por los autores como capaces de proveer métodos para el desarrollo de disciplinadas prácticas de observación, examen y descripción de la experiencia consciente. Puede pues entenderse el programa neurofenomenológico como un segmento del enfoque enactivo por cuanto la apuntada insistencia en el valor para el desarrollo de una ciencia cognitiva con los pies en la tierra del recabado de informes fenomenológicos rigurosos aparece en *The Embodied Mind* como uno de los aspectos del referido enfoque enactivo, un aspecto posteriormente elaborado de forma explícita como una metodología para el estudio científico de la conciencia: la neurofenomenología (Varela, 1996; 1997; 1999).

En un seminal artículo publicado en *Journal of Consciousness Studies* cinco años después de la aparición de *The Embodied Mind*, Varela definía la neurofenomenología como sigue: "Neuro-phenomenology is the name I am using here to designate a quest to marry modern cognitive science and a *disciplined approach* to human experience, thus placing myself in the lineage of the continental tradition of phenomenology. My claim is that the so-called hard problem (…) can only be addressed productively by gathering a research community armed with new pragmatic tools enabling them to develop a science of consciousness" (Varela 1996: 330).[3] En este mismo artículo describiría también Varela a la neurofenomenología como un proyecto destinado a cerrar la brecha explicativa. Ya en este primer momento se hace evidente, por una parte, la atención que la neurofenomenología se propone prestar tanto a los problemas científicos como a los filosóficos abordados en los *Consciousness Studies* y, por otra, el protagonismo que la misma concede a la adecuada teorización de la intersección entre las aproximaciones descriptivas a la experiencia consciente desarrolladas en el seno de tradiciones específicas (dentro de la literatura neurofenomenológica, como su propio nombre pone ya de manifiesto, la fenomenología husserliana ha venido jugando un papel destacado) y las técnicas de observación y marcos teóricos dinámicos destinados a aprehender el maremágnum de la actividad nerviosa implicada en la experiencia consciente.

Como indicábamos, el programa metodológico de la neurofenomenología destina esfuerzos a tareas que muchos filósofos contemporáneos considerarían abocados al fracaso, pues se propone abordar el problema duro y cerrar la brecha explicativa. Pero, ¿cómo pretende hacerlo? Mostrando, respondería el neurofenomenólogo, que las propiedades fenoménicas de los estados mentales conscientes pueden ser integradas en

[3] Cursivas en el original.

marcos explicativos en los que aparezcan en continuidad con las aceptadas por las ciencias naturales (Roy et al., 1999). Desde el punto de vista de los partidarios de este programa metodológico, este objetivo dependería, por una parte, del desarrollo de una rigurosa metodología de observación y descripción en primera persona de la experiencia consciente (de aquí la fenomenología de la neurofenomenología) y, por otra, del de los apropiados modelos explicativos de la dinámica neuronal. Además, se haría necesaria la elaboración de modelos capaces de vincular éstos con aquélla, es decir, modelos capaces de trazar *vínculos significativos* (Varela, 1996: 340) entre dichas descripciones fenomenológicas y los adecuados modelos del funcionamiento nervioso. Como subrayaremos más adelante, las constricciones relativas al modo de trazar estos puentes correrían, según los neurofenomenólogos, en ambas direcciones (Varela, 1996; Varela & Shear, 1999), esto es, de las descripciones fenomenológicas a los modelos explicativos de la actividad nerviosa y viceversa.

El programa neurofenomenológico descansa, pues, sobre dos pilares: el desarrollo de métodos observacionales y marcos descriptivos adecuados de la experiencia consciente, por una parte, y el trazado de vínculos entre éstos y modelos teóricos dinámicos destinados a dar cuenta de la actividad nerviosa implicada en la experiencia consciente, por otra. Nos ocuparemos de estos dos extremos por separado en los siguientes apartados.

## 1._El problema descriptivo

Respecto del primer punto, la fenomenología de la neurofenomenología ha venido incorporando influencias provenientes de tradiciones tanto orientales como occidentales, pero, con todo, y tal y como cabría de hecho esperar dado el propio nombre de este programa metodológico, la preponderancia de la tradición fenomenológica europea no ha dejado de notarse, especialmente en el aparato descriptivo utilizado por la mayoría de los partidarios de dicho programa. No obstante, y a pesar de la aludida pluralidad de influencias, algunos neurofenomenólogos han tratado de ofrecer una definición amplía de aquello a lo que se refieren cuando hablan de fenomenología:

> Phenomenology in th[e] broad sense can be understood as the project of providing a disciplined characterization of the phenomenal invariants of lived experience in all of its multifarious forms. By 'lived experience' we mean experiences as they are lived and verbally articulated in the first-person, whether it be lived experiences of perception, action, memory, mental imagery, emotion, attention, empathy, self-consciousness,

> contemplative states, dreaming, and so forth. By 'phenomenal invariants' we mean categorical features of experience that are phenomenologically describable both across and within the various forms of lived experience. By 'disciplined characterization' we mean a phenomenological mapping of experience grounded on the use of 'first-person methods' for increasing one's sensitivity to one's own lived experience (Lutz & Thompson, 2003: 32; Thompson, Lutz & Cosmelli, 2005: 45-46).

Los partidarios del enfoque neurofenomenológico, no obstante –esto es, nuevamente a pesar de la notable variedad de escuelas introspectivas que han ejercido su influjo sobre la fenomenología de la neurofenomenología–, han venido insistiendo en la adecuación del método de observación y descripción de la experiencia consciente desarrollado en la tradición husserliana, distinguiéndolo radicalmente de la "mera" introspección (Varela, 1996: 338; Thompson, Noë y Pessoa, 1999: 571). En ese método descansa principalmente la esperanza neurofenomenológica de vencer la conclusión a la que arriba Eric Schwitzgebel en *Perplexities of Consciousness*, la de que estamos mejor equipados para el estudio del mundo exterior que para el de la conciencia (Schwitzgebel, 2011: 159). Con todo, describir, como hacen Roy y colaboradores (Roy et al., 1999: 18), la fenomenología husserliana como una empresa dedicada al cuidadoso establecimiento, a través de la descripción y el análisis en primera persona, de datos fenomenológicos entendidos como aquello de lo que realmente somos conscientes por contraposición a aquello que tendemos a creer que habita nuestros mundos fenomenológicos, no parece servir para establecer sólidas demarcaciones entre los métodos fenomenológicos propugnados por los neurofenomenólogos y la "mera" introspección interpretada, poco más o menos, en cualquier sentido que quiera darse al término (desde Titchener o Wuzburgo hasta el uso ordinario del mismo, que cabe despachar mediante su entrada, de una sola acepción, en la vigésima segunda edición del diccionario de la lengua española de la RAE: "Observación interior de los propios actos o estados de ánimo o de conciencia").

Tratando, en cualquier caso, de establecer una metodología fenomenológica, Varela y Shear han subrayado que "lo subjetivo está intrínsicamente abierto a la validación intersubjetiva" (Varela & Shear, 1999: 2τ) y que sólo nos es necesario a tal fin proveernos de un método y un procedimiento para recoger los frutos de esa apertura. Así, el proyecto neurofenomenológico depende, en primera instancia, de la posibilidad del establecimiento de una serie de procedimientos de obtención de datos fenomenológicos que puedan ser tenidos por fidedignos y de los que pudiera servirse la comunidad científica en su avance hacia una ciencia de la conciencia. La idea de que ésa es una posibilidad abierta no se basa, según los proponentes de este programa metodológico, ni en la

asunción de una supuesta infalibilidad o incorregibilidad por parte de los sujetos experimentales respecto de sus propias vidas mentales (Lutz & Thompson, 2003: 39), ni en la de un supuesto *acceso privilegiado* a un mundo fenomenológico estable, acabado, cerrado, inmodificable o apodíctico, pues entienden que el carácter dinámico e inmediato del darse de los fenómenos subjetivos no debiera confundirse con "las claves de su forma de *constitución* y *evaluación*" (Varela & Shear, 1999: 2τ).[4] Por otra parte, una de las asunciones de las que, en efecto, parte el programa neurofenomenológico es la de la irreductibilidad de la experiencia consciente, dado que los datos fenoménicos, dirán Varela y Shear (Ibíd.: 4), no pueden reducirse ni derivarse desde la perspectiva de la tercera persona. Una última asunción que resulta interesante subrayar en este punto es la de que la demarcación entre lo mental accesible a la experiencia y lo mental subpersonal es flexible. Este supuesto juega un papel destacado en el marco de la metodología neurofenomenológica en cuanto aparece relacionado con las posibilidades que, haciendo caso de la literatura producida en el seno de esta joven tradición, abre la práctica prolongada de una observación cuidadosa de la propia experiencia.

Desde el punto de vista neurofenomenológico se hace necesario, como venimos señalando, el establecimiento de una metodología de observación y recogida de datos fenomenológicos que habrá en cualquier caso de partir, como expondremos brevemente más adelante, de una superación inicial por parte del sujeto del punto de vista cotidiano sobre la experiencia consciente, y en este sentido insisten en que la actitud de *familiaridad* con que nos relacionamos con nuestras experiencias conscientes debe subordinarse a un cuidadoso y disciplinado examen de aquello que realmente pulula por nuestros mundos fenomenológicos de cara a producir descripciones ricas y sutilmente interconectadas. Este proyecto, que así expuesto y desde determinado punto de vista puede sonar incluso disparatado –pues, dirían algunos, como Searle (1989a: 203; 1992: 122 del original, 131 de la traducción; 1997a: 112 del original, 118 de la traducción), por lo que a la experiencia consciente toca, la realidad es la propia apariencia–, tendría a su base, a modo de instancias garantes de éxito, las denominadas "metodologías en primera persona", un abigarrado conjunto de continuadas prácticas disciplinadas (presentes tanto en la tradición fenomenológica y en diversas corrientes orientales de meditación como en la psicoterapia) de las que servirse de cara a incrementar y afinar la sensibilidad hacia los propios estados mentales conscientes en diversos planos temporales me-

[4] Cursivas en el original.

diante un aplicado trabajo introspectivo destinado a depurar, por ejemplo, aspectos atencionales o emocionales aplicables a cualquier experiencia consciente. La necesidad que en este punto encuentran los neurofenomenólogos de introducir estas metodologías en primera persona arraiga en el supuesto según el cual existe una poco discutible variabilidad en la capacidad de los sujetos como observadores e informadores de sus propias experiencias, sumado al de que dicha capacidad puede mejorar mediante dichas metodologías. Así las cosas, no parece descabellado plantear la pregunta acerca de los efectos sobre la propia experiencia de los sujetos del entrenamiento en estas disciplinadas prácticas de observación y descripción de la experiencia de las que constan las denominadas metodologías en primera persona ¿Sigue el sujeto, tras años de entrenamiento, teniendo las mismas experiencias sólo que siendo ahora capaz de aprehenderlas e informar de ellas con mayor solvencia, o son las propias experiencias, antes que sus capacidades de observación y descripción, las que han variado? No pretendemos que exista una respuesta correcta a esta pregunta, pero lo cierto es que no cabe afirmar que desde las filas neurofenomenológicas se haya ofrecido una convincente.

Si bien la tradición fenomenológica se presenta como fundamental dentro del marco de las metodologías en primera persona, lo cierto es que la atención a los métodos introspectivos de la psicología científica y los propios de tradiciones orientales (como la meditación samatha, la meditación mahamudra o la meditación trascendental) puede apreciarse en los textos neurofenomenológicos desde los orígenes de esta tradición.

Los métodos de primera persona requeridos por la neurofenomenología habrían de proveer, en cualquier caso, un procedimiento explícito para el acceso a los distintos atributos de sus vidas fenomenológicas, así como una serie de medios efectivos para la expresión y validación de éstos dentro de una comunidad de observadores familiarizados con el procedimiento que sea el caso (Varela & Shear, 1999: 6). A pesar de la heterogeneidad de prácticas de observación y protocolos de comunicación de la experiencia habida en el conjunto de las diferentes tradiciones introspectivas contempladas por los neurofenomenólogos, cabe extraer algunos elementos metodológicos comunes a las mismas en los que la neurofenomenología, en último término, encontraría el fundamento de su confianza en la posibilidad de obtener datos válidos mediante informes de experiencia consciente en primera persona. El primer elemento común que Varela y Shear destacan (Ibíd.: 7-8) es la ἐποχή o reducción fenomenológica (aunque sin mencionarla explícitamente en el fragmento al que aludimos, sino sólo haciendo mención de sus

rasgos definitorios): cuando en un marco experimental neurofenomenológico al sujeto se le pide que realice una determinada tarea, la experiencia concurrente puede ser entendida como un "contenido vivido" primario, del que no saldríamos en la actitud natural (el trato que habitualmente mantenemos en la vida cotidiana con nuestra experiencia consciente, considerándola evidente, inmediata y positivamente disponible, aunque como un supuesto no explícitamente atendido), pero el cual puede ser objeto, tras una suspensión de dicha actitud natural, a su vez, de un segundo estado producto del acto de apuntar al primero en un intento de captarlo para examinarlo, describirlo, etc. Como resultado de este acto, el estado mental cuyo objeto sería el primer estado irreflejo presentará características distintivas que, cabe entender, podrían a su vez ser objeto de sucesivos estados dados por sucesivos actos de vuelta sobre la experiencia. Con independencia de la imputabilidad de este *recursus ad infinitum*, para que surja esta relación entre un estado objeto y otro estado producto del acto de hacer del primero su objeto, debe darse la aludida suspensión de la actitud natural. A este primer elemento común de la reducción fenomenológica aparece asociado un segundo elemento común a las tradiciones introspectivas contempladas por los neurofenomenólogos: la distinción entre contenidos y procesos (a través de los cuales aparecen aquéllos). Estas diferentes tradiciones atienden de formas diversas a esta distinción, destacando el primer término de la misma sobre el segundo en tanto objeto de su método introspectivo (como es el caso de la tradición fenomenológica o la psicología introspeccionista de finales del siglo XIX y principios del XX), o viceversa (como es el caso de diferentes escuelas de meditación). La suspensión de la actitud natural conduce en cualquier caso a un trato con y a una comprensión de los procesos del todo ausente cuando el núcleo atencional, el "centro" de la experiencia lo ocupa, como en la actitud natural, el contenido.

La reducción fenomenológica puede entenderse, en pocas palabras, como un intento de atender a lo que realmente se da en la experiencia tal y como es directamente vivida sorteando el escurridizo óbice de lo que habitualmente suponemos que ella es o en ella se muestra, como una práctica de atención reflexiva a la experiencia consciente que intensifica aspectos tácitos o de trasfondo de ésta –fuera del alcance de la actitud natural– mediante la inducción de estados de atenta autoconciencia que bloquean las creencias y los hábitos arraigados de juicio, explicación y descripción de la experiencia consciente, permitiendo la emergencia de una actitud descriptiva desprejuiciada. Partiendo de esta suspensión del juicio, la atención del sujeto puede desplazarse desde la unilateral atención *natural* al noema o contenido de la experiencia hacia los aspectos

cualitativos pre-reflexivos del proceso experiencial. Dados estos pasos iniciales de suspensión y redirección de la atención, asumen los neurofenomenólogos, el campo de la experiencia subjetiva puede verse progresivamente ampliado y refinado mediante una práctica continuada: la atención alcanzaría así aspectos de la experiencia consciente habitualmente desatendidos.[5] Este ciclo de suspensión, redirección y ampliación, mediado y fomentado por la práctica sostenida, proporcionaría la posibilidad de obtener informes fenomenológicos validables intersubjetivamente, de suerte que el paso crucial de las metodologías en primera persona, el de la comunicación y descripción de la experiencia consciente, constituye el paso decisivo en el tránsito hacia la intersubjetividad de los datos subjetivos o de primera persona, un paso en virtud del cual se haría posible calibrar y relacionar estos datos subjetivos con datos objetivos o de tercera persona (Lutz & Thompson, 2003: 38). Con este breve esquema[6] de la reducción fenomenológica, obviamente, gran cantidad de extremos y técnicas propias tanto de la tradición fenomenológica como de otras tradiciones introspectivas permanecen intocados. No obstante, con él, el núcleo de los aspectos compartidos por estas diversas tradiciones tampoco resulta particularmente violentado –cabe no obstante hacer soslayada mención en este punto, por una parte, de la existencia de una tendencia actual a la búsqueda de intersecciones y contrapesos entre diferentes tradiciones y enfoques, y, por otra, de la creciente insistencia en las metodologías de primera persona como algo encarnado y, decisivamente, *pragmático* (Depraz, Varela & Vermersch, 2003), por contraposición al enfoque *teórico* de la fenomenología denunciado en el segundo capítulo de *The Embodied Mind* (Varela, Thompson & Rosch, 1991).

Otro elemento común a las diversas tradiciones introspectivas mencionadas al que hemos aludido ya, por más que tangencialmente, consiste en la idea de que los sujetos que practican las técnicas de introspección propias de cada una de ellas alcanzan pericia progresivamente. Así, entienden los neurofenomenólogos, una práctica continuada de suspensión de la actitud natural resulta necesaria para alcanzar aprehensiones y descripciones más ricas de la experiencia. No es, pues, de extrañar que la neurofenomenología se nutra de la idea según la cual en la observación de la propia experiencia hay más de lo que parece haber a simple vista (Varela & Shear, 1999: 13) –y alcanzar a observar y

[5] Cabe señalar que en este punto los neurofenomenólogos se ven enfrentados a un problema inverso pero análogo al de los psicólogos de la atención a la hora de definir la noción de atención sin echar mano de la de conciencia (para una revisión crítica de este extremo, consultar el último capítulo de Styles, 2006).

[6] Un tanto unilateral, pues lo hemos resuelto haciendo compendiosa y unilateralmente uso de nociones y planteamientos explícitamente fenomenológicos y poco menos que obviando totalmente las propuestas importadas por los neurofenomenólogos desde otras tradiciones.

describir debidamente ese más es el objetivo de la aludida práctica continuada de suspensión de la actitud natural–. Otro elemento común, en mayor o menor medida, a estas tradiciones introspectivas es el del papel de la segunda persona entendida como *otro* presente y activo en el proceso de acendramiento de la aprehensión fenomenológica. Este otro, esta segunda persona, en la posición de mediador de algún modo presente en el proceso de la práctica continuada, ofrecería valiosas indicaciones del mismo modo en que lo hace un entrenador deportivo: no participa en las competiciones sino indirectamente, pero nadie duda de la importancia de su papel. Esta segunda persona, además de ofrecer pistas útiles durante el proceso de la práctica introspectiva continuada, sirve para establecer una posición de validación de los datos de primera persona diversa de la habitual posición de validación de la tercera persona que el experimentador científico ocupa, dado que el sujeto, la primera persona, y la señalada segunda persona comparten tanto un campo común de experiencia –entre otras cosas, y ante todo, a causa de su conespecifidad– como unos medios descriptivos en los que ambos se hallan análogamente inmersos.

En definitiva, los métodos de primera persona prohijados por los neurofenomenólogos sirven a dos objetivos considerados cruciales por los proponentes de esta metodología: aumentar la receptividad o sensibilidad hacia la propia experiencia, y pergeñar un medio para la obtención de informes de la experiencia consciente que puedan ser tenidos por válidos o intersubjetivamente validables. Ambos objetivos se hallan vinculados, pues una de las más importantes asunciones dentro de la metodología neurofenomenológica consiste en la idea de que las prácticas sostenidas de observación sistemática de la propia experiencia permiten que segmentos pre-reflexivos y pre-lingüísticos de la misma acaben estando subjetivamente disponibles y puedan ser descritos con una precisión creciente, apoyada además en la presencia de segundas personas familiarizadas con las prácticas introspectivas y medios descriptivos que sean el caso.

En esta serie de prácticas y supuestos, en cualquier caso, descansa la confianza depositada por los neurofenomenólogos en la posibilidad de obtener datos de experiencias subjetivas con un estatus tan poco dudoso desde el punto de vista científico como los obtenidos mediante cualesquiera medios experimentales de tercera persona. En este sentido, Lutz y Thompson han llegado a proponer que los datos obtenidos utilizando métodos de primera persona mediante los cuales se haga dable la posibilidad de explicitar y describir lo de otro modo tácita e inadvertidamente presente en la experiencia pue-

den de hecho entenderse como "disponibles para una caracterización intersubjetiva y objetiva" (Lutz & Thompson, 2003: 37τ).

Un destacado aspecto de la aproximación neurofenomenológica al estudio en primera persona de la conciencia que es necesario mencionar antes de pasar al siguiente apartado es el de, por así decir, su concepción global de la conciencia, asociada a una serie de categorías características tomadas de la tradición fenomenológica que, sin añadir determinados matices, no acaban de encajar con los términos en los que habitualmente se discute el problema de la conciencia en el ámbito "analítico" anglosajón. Así, mientras en el ámbito de la filosofía anglosajona ha venido destacando la distinción entre la conciencia fenoménica y la conciencia representacional (Villanueva, 1995: 391) o la conciencia de acceso (Block, 1995; 2001; 2003; 2007), en el ámbito neurofenomenológico la cuestión de las formas de conciencia se plantea de otro modo: son aquí cruciales las nociones de intencionalidad, transitividad/intransitividad y autoconciencia pre-reflexiva. Esta última noción ocupa un lugar central en el marco metodológico de la neurofenomenología y refiere a una forma primitiva de ipseidad, más bien, de autoconciencia: al sentimiento crudo de ser uno que resulta indisociable de cualquier experiencia consciente. De este modo, cualquier experiencia consciente dotada de un determinado contenido intencional, estará al tiempo dotada de esta forma de autoconciencia: toda experiencia de este tipo, además de apuntar a su objeto (conciencia transitiva), se hace de algún modo manifiesta a sí misma (conciencia intransitiva) de forma involuntaria y simultanea a la conciencia *del* objeto sin necesidad de ninguna clase acto subsecuente de introspección o reflexión voluntaria sobre la experiencia, y sin formación alguna de juicios o creencias. Puede decirse que este tipo de autoconciencia está dotada de una reflexividad intrínseca, involuntaria y automática (hay en ella un darse a sí misma), pero es irreflexiva en el sentido de que aparece sin necesidad de ninguna clase de acto explícito de reflexión. Esta distinción aparece en el ámbito neurofenomenológico, por así decir, en paralelo a la distinción –clásica en la tradición fenomenológica– entre el proceso noético de experienciar y el objeto noemático de la experiencia: en términos de esta distinción clásica, toda experiencia intencional comportará además de conciencia de su noema, una tácita conciencia noética de sí misma que, tanto en la tradición fenomenológica como en la breve tradición neurofenomenológica ha sido vinculada a la subjetividad encarnada en un cuerpo (Merleau-Ponty, 1945; Varela, Thompson & Rosch, 1991; Zahavi, 2002). Desde este punto de vista, la distinción anglosajona entre

fenomenalidad y acceso adquiere matices diversos de los habituales en la literatura sobre filosofía de la mente y ciencias cognitivas, matices provenientes de la concepción dinámica de la estructura noética-noemática de la conciencia sostenida por fenomenólogos y neurofenomenólogos. Así, desde el punto de vista fenomenológico adoptado por la neurofenomenología, toda *experiencia vivida*[7] se caracteriza por la presencia de eventos y procesos metales pre-verbales y pre-reflexivos dotados de valores emocionales que, aunque no resulten inmediatamente accesibles al pensamiento o el lenguaje, son intransitivamente experienciados y tienen un determinado carácter fenoménico asociado a una forma primitiva de auto-conciencia ínsita en toda experiencia y dada la cual, debe entenderse, toda forma de experiencia es en principio accesible, en el sentido de que es la clase de estado que puede presentarse al sujeto como disponible –particularmente haciendo uso de los señalados métodos de primera persona–, es decir, como objeto posible de la conciencia reflexiva, y pasible, por tanto, de convertirse en tema de un informe verbal.

## 2._El problema del enlace

Este segundo punto refiere a la necesidad en que se encontraría el neurofenomenólogo, de cara a cerrar, como se propone, la brecha explicativa, de *vincular significativamente* (Varela, 1996: 340) los informes en primera persona obtenidos gracias a las metodologías en primera persona a las que aludíamos en el apartado anterior con los apropiados modelos neurocientíficos del funcionamiento neuronal, que, desde el punto de vista neurofenomenológico, provendrían de un enfoque corporeizado y dinámico de las bases biológicas de la conciencia (Varela, 1995; Thompson & Varela, 2001; Lutz & Thompson, 2003; Varela & Thompson, 2003). El abordaje neurofenomenológico de este problema que hemos denominado aquí del enlace (no confundir con el *bindig problem*) está marcado por las constricciones bidireccionales a las que aludíamos escuetamente más arriba y que constituyen el núcleo de la que Varela denominara *hipótesis de trabajo de la neurofenomenología*, en la cual se establece el fundamento de dichas constricciones bidireccionales mediante la *metodología de constricciones recíprocas*

[7] Traducimos aquí "lived experience", una locución muy habitual en la literatura neurofenomenológica y que se refiere a la experiencia subjetiva de la actividad mental tal y como es vivida y verbalmente articulable (Lutz & Thompson, 2003: 32; Thompson, Lutz & Cosmelli, 2005: 45-46). "Lived experience is where we [neurophenomenologists] start from and where we all must link back to, like a guiding thread" (Varela, 1996: 334).

(Varela 1996: 343; van Gelder 1999b: 246). Dicha propuesta metodológica se basa en el supuesto de que las referencias fenomenológicas a la estructura de la experiencia y sus equivalentes en datos neurocientíficos de tercera persona se relacionan unas con otras a través de restricciones mutuas: la consideración conjunta del plano biológico y el fenomenológico de lo mental habría de ocasionar u ofrecer en el plano metodológico la posibilidad de establecer limitaciones y validaciones mutuas. Esta es la idea que plantea la hipótesis de trabajo de la neurofenomenología y que en un plano heurístico puede ser interpretada en los siguientes términos: datos provenientes de rigurosas descripciones fenomenológicas pueden servir de guía u orientación a la hora de trabajar con datos acerca de la neurodinámica a gran escala de la conciencia (para describirlos, analizarlos, ponderarlos, interpretarlos, discutirlos), y viceversa (vid., v. g., Lutz et al., 2002). En este sentido, la metodología neurofenomenológica enfatiza la utilidad de trabajar con minuciosos informes fenomenológicos para el avance hacia una ciencia de la conciencia y, particularmente, subraya la importancia de trabajar con informes fenomenológicos provistos por sujetos experimentales entrenados en metodologías de primera persona de cara a especificar los procesos fisiológicos relevantes para la conciencia. Se haría así, desde la perspectiva neurofenomenológica, necesaria la obtención de ricas descripciones fenomenológicas provistas por el disciplinado examen fenomenológico de la experiencia consciente llevado a cabo por sujetos entrenados en metodologías de primera persona, y, al tiempo, la utilización de dichas descripciones en el acotamiento de modelos explicativos e interpretaciones adecuadas de los registros de procesos fisiológicos implicados en la experiencia consciente. La propuesta metodológica neurofenomenológica trataría así de proporcionar la posibilidad de ampliar el rango de los procedimientos y protocolos que suelen contemplarse en programas de investigación neurobiológica de la conciencia mediante la introducción de rigurosos métodos de examen y descripción fenomenológicos, así como la de generar nuevos datos a partir de las indicaciones proporcionadas por dichos exámenes y descripciones, esto es, haciendo uso de la luz que minuciosos informes fenomenológicos puedan arrojar sobre la investigación de las bases biológicas de la conciencia.

Por lo que toca a la otra dirección de la metodología de constricciones recíprocas, como puede intuirse ya, según la hipótesis de trabajo de la neurofenemenología los datos de tercera persona obtenidos en el contexto de protocolos experimentales neurobiológicos habrían asimismo de poder ofrecer a los sujetos experimentales valiosas indicaciones acerca de aspectos de sus mundos fenomenológicos que les hubieran podido pa-

sar de otro modo inadvertidos (vid. Le Van Quyen & Petitmengin, 2002). La metodología neurofenomenológica pretende en definitiva ofrecer en el marco del actual camino hacia la elaboración de una ciencia de la conciencia, la posibilidad de recabar y articular teóricamente datos interrelacionados de primera y tercera persona.

La propuesta metodológica neurofenomenológica de las constricciones recíprocas, señalemos de pasada, ha sido también desarrollada en términos causales. Así, por ejemplo, Varela y Thompson plantean que existen dos tipos de causación en el cerebro, una ascendente (*upward*), o de lo local a lo global, y otra descendente (*downward*), o de lo global a lo local, e interpretan este planteamiento en términos de efectiva existencia de dos direcciones causales entre lo fenoménico y lo neuronal, identificando el tipo descendente de causación con una forma de causación que iría de lo fenoménico a lo neural, en tanto que lo fenoménico es concebido como un parámetro de la dinámica de agrupaciones neuronales de gran escala, dándose así, según los autores, una serie de relaciones causales y explicativas recíprocas (Thompson & Varela, 2001: 421).

Aludíamos tangencialmente en las primeras líneas del párrafo con el que abríamos este segundo apartado a la apuesta neurofenomenológica según la cual la mejor opción a la hora de seguir la pista de los así llamados correlatos neuronales de la conciencia (NCC) será un enfoque basado en la teoría de sistemas dinámicos concretado en modelos formales destinados a capturar la dinámica de las relaciones entre grandes asambleas neuronales ampliamente distribuidas. El problema de vincular fenomenología y biología abarcaría desde el punto de vista de esta apuesta, por una parte, y como señalábamos en el apartado anterior, el desarrollo de refinadas técnicas de observación y descripción fenomenológica, y, por otra, la búsqueda de una base neurofisiológica capaz de jugar su papel en el marco de la metodología de constricciones recíprocas, una búsqueda orientada para los partidarios del enfoque neurofenomenológico, en el plano más general, por el enfoque corporeizado y enactivo en ciencias cognitivas (el cual, por otra parte, sanciona como insuficiente la consideración de las bases biológicas de la conciencia como algo meramente neuronal, a pesar del incuestionablemente privilegiado papel que las neurociencias desempeñan dentro del programa metodológico de la neurofenomenología), cuya manifestación metodológica en el contexto de la vía neurofenomenológica hacia una neurobiología de la conciencia consistiría en herramientas analíticas y modelos formales para la descripción de los procesos fisiológicos relevantes para la conciencia que la teoría de sistemas dinámicos ofrece. Dichos procesos, centrándonos en el

plano *meramente*[8] neuronal, y tal como actualmente es ampliamente admitido (vid., v. g., Tononi & Edelman, 1998; Freeman, 1999; Edelman & Tononi, 2000; Dehaene & Naccache, 2001; Engel & Singer, 2001; Seth & Edelman, 2009), consistirían en una constante orquestación transitoria de agrupaciones neuronales dispersas y funcionalmente especializadas, una orquestación que, en cada instante, vincularía asambleas neuronales ampliamente distribuidas que, en tanto coaliciones interinas, resultarían desde el punto de vista de su actividad fisiológica altamente integradas al tiempo que elevadamente diferenciadas, y que estarían conectadas por bucles de reentrada, esto es, por conexiones recíprocas. El comportamiento de estas coaliciones neuronales ampliamente distribuidas resulta actualmente difícil de conceptualizar y de subsumir en modelos formales:[9] los principios de su funcionamiento, sus propiedades específicas y los mecanismos causales que determinan el comportamiento dinámico global de los procesos cerebrales a gran escala son el objeto de la reciente línea de investigación acerca del "problema de la integración a gran escala" (Varela et al., 2001), una línea de investigación cuyos modelos en desarrollo han de lidiar con las escurridizas propiedades de unas agrupaciones neuronales complejas, caracterizadas por procesos dinámicos que tienen lugar en el plano temporal de los milisegundos y que presentan formas de comportamiento endógenas y autoorganizadas muy variables de ensayo a ensayo, motivo por el cual resultan difícilmente controlables experimentalmente. Nuevamente surgen requerimientos duales difícilmente abordables a día de hoy: se hace necesario, por una parte, el desarrollo de un marco conceptual y teórico adecuado para la comprensión de la complejidad neuronal implicada en las dinámicas neuronales concebidas como cruciales para la experiencia consciente y, por otra, encontrar el modo de relacionar de forma explícita y rigurosa ese marco teórico en desarrollo con informes fenomenológicos fidedignos. Veamos por separado el modo en que se plantea el abordaje de ambos desafíos desde el punto de vista de la metodología neurofenomenológica.

[8]El uso de este adverbio en este punto pretende destacar la laxitud del compromismo de los neurofenomenólogos con el enfoque corporeizado y enactivo que reivindican: a pesar de subrayar ese compromismo, a la hora de la verdad, toda actividad biológica a excepción de la nerviosa es desatendida en los estudios neurofenomenológicos.

[9] Vid., no obstante, Le Van Quyen (2003).

## 2.1._Dinámica neuronal

El marco teórico de la teoría de sistemas dinámicos complejos se presenta, desde el punto de vista neurofenomenológico,[10] como crucial por lo que al intento de hacer frente al primero de los señalados desafíos o requerimientos se refiere (Kelso, 1995; Freeman, 2000b; Thompson & Varela, 2001; Varela et al., 2001; Le Van Quyen & Petitmengin, 2002; Le Van Quyen, 2003; Thompson, Lutz & Cosmelli, 2005: 42 y ss.; Thompson, 2007; Le Van Quyen, 2010). El desarrollo de un marco conceptual y teórico adecuado para la comprensión de la compleja actividad neuronal implicada en las dinámicas fisiológicas cruciales para la experiencia consciente, entienden los proponentes de la metodología neurofenomenológica, habrá de atender, para alcanzar a esclarecer los mecanismos de la integración nerviosa a gran escala concebida como determinante de la emergencia de la conciencia, antes que a las propiedades, la actividad o las funciones particulares de los diversos sistemas neuronales funcionalmente especializados que pudieran integrar, en términos de Edelman y Tononi (Edelman & Tononi, 2000), un *núcleo dinámico* concreto, a la naturaleza dinámica de las conexiones recíprocas entre las mismas. Guardando las distancias, cabe hablar aquí del acuerdo al que los partidarios de este enfoque dinámico llegarían con Searle al desestimar de consuno las posibilidades de una aproximación a las bases biológicas de la conciencia partiendo de una investigación circunscrita al nivel de los circuitos especializados, los núcleos específicos, las clases de neuronas o los neurotransmisores predominantemente asociados a una determinada clase de éstas, una aproximación que Searle ha denominado estrategia *building block* (Faigenbaum, 2003: 51; Searle, 2000: 53 y ss.; 2007: 175).[11] Se buscan, pues, variables capaces de dar cuenta de la emergencia y el cambio en los patrones de integración de la actividad nerviosa a gran escala. Este tipo de variables han tratado de definirse, por ejemplo, tomando como referencia la sincronización de fase de grupos neuronales ampliamente distribuidos, describiendo y cuantificando patrones transitorios de acti-

[10] El vínculo entre la teoría de sistemas dinámicos y la fenomenología se establece a partir de la idea de que la intencionalidad constitutiva de un sistema cognitivo corresponde a una forma de autoorganización (vid., v. g., Thompson, 2007: 27).

[11] Hablábamos al comienzo de esta larga frase de unas distancias a guardar que, aclaremos, no tendrían sino que ver con la vaguedad con la que –en comparación con quienes advocan por el señalado enfoque dinámico– Searle plantea su crítica a lo que denomina orientación "building block" en la investigación neurocientífica de la conciencia.

vidad neuronal oscilante en diversas regiones encefálicas y diversas frecuencias (Rodriguez, et al., 1999).[12]

Gran cantidad de investigación básica y desarrollos teóricos se hacen necesarios para concretar estas propuestas, pero, a pesar de ello, tanto desde dentro de las filas neurofenomenológicas como desde fuera de las mismas la apuesta general gana pujanza y puede considerarse como el paradigma actual para una neurociencia de la conciencia que más respaldo ha venido obteniendo dadas no sólo las evidencias experimentales recabadas, sino asimismo la capacidad predictiva de los marcos teóricos desarrollados dentro del mismo. Dicha apuesta, para repetirlo, consistiría, en su esqueleto elemental, en la asunción según la cual una actividad neuronal distribuida a gran escala, esto es, no circunscrita a regiones o circuitos específicos, y dotada de coherencia y una forma recursiva de interacción es el mejor de los candidatos para las bases fisiológicas de la conciencia. No obstante, los detalles, en ningún caso insignificantes, varían ostensiblemente de unas propuestas a otras. Una asunción habitual dentro de este marco paradigmático aceptada de buen grado por los partidarios del enfoque neurofenomenológico es la de que esa coherencia a gran escala se halla estrechamente vinculada con procesos de autorregulación determinantes para la emergencia la integración neuronal que, tal y como ha venido poniéndose de manifiesto en las últimas décadas, resulta fundamental de cara a posibilitar la experiencia consciente. Las propiedades dinámicas de esta integración y los principios de su modo de funcionamiento a diferentes escalas temporales y espaciales requieren por el momento de gran cantidad de investigación y elaboración teórica y formal (matemática) para ser esclarecidas.

Hemos dejado caer en varias ocasiones que el marco metodológico de la neurofenomenología apunta más allá de las bases neurobiológicas de la conciencia a causa de su filiación enactiva, situada y encarnada. Así, desde el punto vista adoptado por los proponentes de este programa metodológico, la necesaria integración de la actividad neuronal de la que venimos hablando no debe ser concebida de un modo aislado, pues

[12] Justifica este recurso a la sincronía de fase como medio para describir los mecanismos de la integración de la actividad nerviosa a gran escala la acumulación de evidencias que apuntan a ella como crucial en procesos de atención, memoria o integración sensomotora (vid., v. g., Varela, 1995; Tononi & Edelman, 1998; Dehaene & Naccache, 2001; Engel & Singer, 2001; Engel, Fries & Singer, 2001; Ward, 2003), además de su probado potencial predictivo, tanto desde el punto de vista conductual, como desde el fisiológico o el experiencial (Engel, Fries & Singer, 2001). Cuando en este contexto se habla de sincronía de fase no se alude, por otra parte, a la mera coherencia temporal de la frecuencia de la actividad nerviosa en diversas áreas encefálicas, esto es, no se trata sencillamente de que la actividad que cupiera captar en diferentes áreas mediante un equipo de electroencefalografía o uno de magnetoencefalografía sea en un momento dado de la misma frecuencia, sino que, adicionalmente, debe la misma hallarse en fase, desplazándose al unísono desde las cretas hacia los valles (vid., v. g., Varela, 2001: fig. 4).

resulta decisivo para una acabada comprensión de la naturaleza de la conciencia encuadrarla teóricamente dentro de un marco en el que aparezca vinculada con el resto del organismo (más allá de los famosos NCC) y con el entorno en el que éste lleva a cabo su actividad, y a este fin ha de conceptualizarse como integrada en los sucesivos ciclos de regulación organísmica, acoplamiento sensomotor organismo-entorno e integración intersubjetiva. De este modo, no sólo la actividad neuronal, sino también la actividad del organismo en su medio se hace necesaria para una adecuada comprensión de la naturaleza de la conciencia, encuadrada en el marco de sucesivos niveles que albergarían procesos acaecidos en y entre el cerebro, el cuerpo y el entorno de los organismos conscientes.

### 2.2._… y experiencia consciente

Una vez detallados de forma precisa estos modelos dinámicos de la neurofisiología de la conciencia, e incluso alguna clase de marco teórico capaz integrar estos modelos con los pertinentes ciclos de actividad encarnada y situada, podemos estar seguros de que muchos filósofos contemporáneos seguirían protestando: "¡sigo intuyendo que entre tus modelos explicativos y tu *explanandum* se abre un abismo!". Es decir, parece que, o bien necesitamos acumular una cantidad de evidencias experimentales y desarrollos teóricos que apenas alcanzamos a imaginar, o bien gran cantidad de contemporáneos nuestros no se equivocan en absoluto cuando afirman que, se ofrezca el *explanans* que se ofrezca dada la apuntada acumulación, el *explanandum* de la experiencia consciente de ningún modo se seguirá de aquél. La postura constructiva –id est, antimisteriana– de la metodología neurofenomenología conduce, claro, a la apuesta por la primera de las opciones. Pero, ¿cómo ponerla en marcha ofreciendo, indirectamente, réplica a los abanderados de la segunda?

La tarea que se impone a la metodología neurofenomenológica desde la perspectiva del avance hacia la posibilidad de ofrecer respuesta a semejante interrogante consistiría en integrar los modelos dinámicos de la actividad neuronal de los que veníamos hablando con los datos fenomenológicos a los que aludíamos más arriba. Pero, ¿cómo integrar estos datos fenomenológicos en esas dinámicas neuronales de un modo explícito y riguroso? Una destacada faceta de la respuesta que el neurofenomenólogo ofrecería a esta pregunta implica una ampliación de protocolos experimentales: haría falta desarrollar marcos experimentales en los que los sujetos, entrenados –al igual que los expe-

rimentadores– en metodologías de primera persona, se encontraran activamente implicados en el proceso de obtención de datos. En un contexto experimental de este tipo, entiende el neurofenomenólogo, los sujetos experimentales pueden contribuir a la descripción e identificación de rasgos fenomenológicos que puedan utilizarse para identificar y describir las propiedades de las dinámicas neuronales de la experiencia consciente. De este modo, cabe imaginar un futuro en el que las evidencias convergentes conduzcan al establecimiento de "jánicas" descripciones dinámicas con potencial predictivo en dos direcciones (feno-neuro y viceversa) mediante el desarrollo de medios para la obtención de informes fenomenológicos lo bastante refinados, precisos y completos como para poder expresarlos en términos dinámicos y formales traducibles, además, a términos análogos referidos a las propiedades de la neurodinámica de la experiencia consciente.

Subrayemos, para concluir con esta compendiosa presentación de los aspectos fundamentales de la propuesta metodológica neurofenomenológica, que ella no sólo afronta un proyecto científico, sino que se enfrenta también a un problema formulado en el ámbito de la filosofía anglosajona de la mente (un ámbito en el que, a pesar de la distancia que cabe entender que media entre él y las filiaciones neurofenomenológicas, se ha venido discutiendo y comentando cada texto neurofenomenológico destacado con gran interés), el de la brecha explicativa, un problema que, según Lutz & Thompson (2003: 32), abarca cuestiones conceptuales, metodológicas y epistemológicas. Quizá se trate, en el caso de este último proyecto (el de arrostrar experimentalmente el problema del *explanatory gap*), de un proyecto más ambicioso y abstracto, dado que en él aparecen, tal y como Lutz y Thompson sugieren, desdibujadas las fronteras entre lo conceptual y lo experimental. En cualquier caso, una vez hallados los *vínculos significativos* que buscan los neurofenomenólogos podemos, nuevamente, estar seguros de que los filósofos de la mente seguirán dividiéndose entre los que opinen y argumenten que la brecha explicativa ha sido salvada con ellos y los que opinen y argumenten lo contrario. En esta situación, cabe preguntar, ¿cómo entiende el neurofenomenólogo que su programa salvará la brecha explicativa? Su respuesta, en congruencia con el proyecto de la ciencia moderna tal y como habitualmente es concebida (una concepción usualmente remontada a Galileo): mediante la matematización. Teniendo en cuenta el rechazo general de aproximaciones funcionalistas por parte de los neurofenomenólogos a causa de su carácter abstracto e ignaro de la contextualidad y la corporeidad de lo mental, puede que resulte para algunos curioso, pero lo cierto es que parecen destinar los partidarios de este programa metodológico importantes esperanzas a una empresa futura de naturaliza-

ción vía formalización de la fenomenología husserliana que permita articularla en modelos fidedignos con "las ciencias naturales relevantes de nivel inferior" (Roy et al., 1999: 63τ). En esto consistiría, en último término, el cardinal frente en desarrollo con el que los neurofenomenólogos intentan salvar la brecha explicativa entre lo fenoménico y lo neuronal: desarrollar modelos matemáticos en los que se hallen implicadas variables que puedan referir tanto a estados fenoménicos como neurofisiológicos, una idea a la que Varela (Varela, 1997) se refería como *pasaje generativo* ("generative passage"), pretendiendo apuntar a un plano neutral respecto de lo fenoménico y lo neuronal en el que tanto lo uno como lo otro pueda ser plasmado.

### 3._Breve comentario crítico

Atravesando con la mirada la profusa colección de tecnicismos fenomenológicos utilizados dentro de la aún joven tradición neurofenomenológica, cabe preguntarse si, efectivamente, difieren las técnicas introspectivas fenomenológicas (no dejemos de advertir que algún fenomenólogo denunciaría un oxímoron en esta locución) de las usualmente utilizadas en los *Consciousness Studies* desde diversos ángulos en protocolos experimentales en neuroimagenología o electroencefalografía. Destaquemos que no somos los únicos en hacernos semejante pregunta: "It seems to me that the methods for collecting first-person data employed by neurophenomenologists are much the same as those employed elsewhere in the study of consciousness" (Bayne, 2004: 355). Bayne pone ejemplos de estudios neurofenomenológicos (Lutz, 2002; Lutz et al., 2002) en los que resulta comprometido afirmar que los sujetos experimentales hacen uso de técnicas fenomenológicas específicas, como la aludida reducción fenomenológica, por cuanto las tareas sólo implican lo habitual en paradigmas experimentales típicos en psicología de la atención o psicología de la percepción sumado a la utilización de un vocabulario fenomenológico específico, cosa que no implica que los sujetos estén haciendo, como pretenden los neurofenomenólogos, nada diferente de la "mera introspección". En vista de la centralidad que los neurofenomenólogos conceden a su metodología introspectiva, debieran evitar a toda costa la mínima la laxitud a la hora de marcar las distancias con la "mera introspección". Según el programa neurofenomenológico, con un método adecuado para la observación y descripción de las experiencias conscientes, salvaremos los escollos con los que nos encontrábamos al trata de validar los informes de primera persona cuando sólo contábamos con la "mera introspección". Siendo esta distinción tan

fundamental para el programa neurofenomenológico como la frase anterior trasluce, parece que, no obstante, no ha sido demasiado lo que los neurofenomenólogos han venido haciendo para trazar los límites entre la "mera introspección" y las prácticas habituales en psicología experimental, por una parte, y la observación y descripción metódica que los partidarios de esta –desde esta perspectiva *supuestamente*– nueva metodología propugnan, por otra. Además, según Bayne, faltaría explicar y explicitar, incluso después de haber sido explícitos en la puesta en práctica de las resultas de la señalada demarcación, los motivos por los cuales los resultados obtenidos haciendo uso de la reducción fenomenológica son más fiables que los de la "mera introspección".

La fenomenología buscaba dejar de lado nuestras teorías, salir de ellas para desde "ahí" ofrecerles fundamento. La neurofenomenología busca algo semejante pero menos ambicioso: pretende brindar la ocasión de mirar por primera vez la mente, desde ninguna parte, pero puede que ello resulte tan hacedero como fundamentar el edificio completo del conocimiento humano desde el indefinido no-lugar al que nos invitara a viajar imaginariamente la epistemología trascendental tradicional. ¿Hay carga teórica en los informes de los sujetos entrenados en metodologías de primera persona? ¿Se halla exenta de teoría la fenomenología de la neurofenomenología? Se trata de un punto ciertamente conflictivo, y Varela admite (Blackmore, 2006: 232) que una de sus formas favoritas de entrenamiento en metodologías de primera persona comporta una teoría de la mente. No podía ser de otra manera, ya que difícilmente cabe esperar una forma de observación científica totalmente libre de cualquier clase de supuesto de trasfondo. En todo caso, y suponiendo que los sujetos entrenados en esta clase de metodologías de primera persona pueden eludir con éxito todas las ideas ingenuas y de corte teórico susceptibles de contaminar los informes en primera persona, debemos aún preguntar: ¿no tenemos que explicar, también, la conciencia tal y como les parece a sujetos ingenuos no entrenados en metodologías de primera persona? Si, por ejemplo, la conciencia perceptiva visual no es un pleno –como muestra, por ejemplo, el fenómeno de la ceguera al cambio–, a pesar de que precisamente así se les presente a la inmensa mayoría de los sujetos, esa apariencia ingenua debe ser explicada por una ciencia de la experiencia consciente, cosa que no cabe hacer partiendo exclusivamente de datos obtenidos de sujetos no ingenuos a los que no se les presente de ese modo. Datos estadísticos provenientes de estudios interculturales pueden ayudar, en una primera aproximación, a dar cuenta de hasta qué punto constituye cada clase de prejuicio *naïve* un fenómeno memético o biológico –nótese que si los sujetos entrenados son inmunes tanto a las precon-

cepciones que nos serían connaturales en tanto conespecíficos como a las influencias culturales y teóricas, sus informes no pueden ayudarnos a diferenciar esas clases, como tampoco pueden por tanto ayudarnos a integrar en un marco teórico solvente las diferencias interculturales en juicios perceptivos del tipo de las que la psicología cultural viene iluminando (vid. Berry, et al., 2002: cap. 8)–. Necesitamos, pues, además de rigurosas metodologías en primera persona y datos fenomenológicos aportados por sujetos entrenados, datos fenomenológicos ingenuos. Una explicación de la tendencia de los sujetos ingenuos a sobreestimar la amplitud y distinción de sus mundos fenomenológicos, o de las diferencias interculturales en percepción o emoción, no sólo requerirá sofisticadas metodologías en primera persona, sino también datos recogidos, por así decir, bajo la perspectiva heterofenomenológica.[13] A lo apuntado cabe añadir una delicada cuestión a la que hemos hecho ya referencia: ¿cambian las metodologías de primera persona sólo los medios de expresión y observación de la experiencia o, asimismo, la propia experiencia?

Por otra parte, una asunción neurofenomenológica que resulta en cualquier caso desconcertante es la de la autonomía e irreductibilidad de dominios que, no obstante, han de restringirse y limitarse mutuamente. Desde un punto de vista conceptual, sin descender al plano ontológico, parece difícil sostener que dos conjuntos de conceptos que no mantienen relaciones implicativas o de cualquier otra clase, es decir, que son netamente autónomos, ofrezcan de algún modo la posibilidad de limitar o restringir sus ámbitos mutuamente. Dos dominios diferenciados y conceptualmente autónomos difícilmente podrán mantener semejante clase de relaciones. El punto al que queremos conducir ahora la atención puede hallarse muy explícitamente en Thompson, Noë & Pessoa (1999: 195), en un fragmento en el que las tensiones entre el compromiso con la autonomía e irreductibilidad de lo fenomenológico y el compromiso con la tesis del equilibrio reflexivo implícito en la metodología de las constricciones mutuas se hacen patentes cuando los autores subrayan que la autonomía conceptual del ámbito fenomenológico viene dada por una clase de autocomprensión como sujetos conscientes que los seres humanos, sin duda, tenemos, y que hemos de diferenciar de nuestra comprensión del plano de los datos de tercera persona, aún cuando, no obstante, ambos planos puedan

[13] No sólo la psicología experimental podrá, por lo tanto, proporcionar datos, constructos y marcos teóricos de interés para los *Consciousness Studies*, sino asimismo la psicología diferencial y la metodología correlacional.

adaptarse mutuamente en un "equilibrio reflexivo". Se nos dice, pues, que es posible vencer la señalada tensión, pero no se nos indica de qué modo.

En relación con la interpretación causal de la metodología de constricciones recíprocas, cabe señalar que se trata de una concepción que, ciertamente, no acaba de casar del todo bien con el antirreduccionismo del que parten los neurofenomenólogos, ni con su concepción corporeizada y situada de los fenómenos mentales. Desde el punto de vista enactivo, cualquier especificación de las bases neurofisiológicas de la conciencia habrá de ser contemplada como algo parcial, pero no puede decirse que la literatura neurofenomenológica ofrezca las claves para convertir el eslogan de la situación y la corporeización en paradigmas experimentales y herramientas heurísticas o interpretativas concretas. Todos los estudios neurofenomenológicos de los que tenemos noticia son estudios en los que lo único que se añade a la investigación neurocientífica usual es un vocabulario específico: la actividad encarnada del organismo sigue donde estaba (desde luego, no dentro de las máquinas de PET o fMRI o los equipos de EEG). Si los correlatos neuronales de la conciencia son descartados como incompletos al no hallarse inscritos en el contexto mayor de la acción del organismo en su entorno, los neurofenomenólogos, o lo han dicho muy bajo, o no nos han dicho cómo buscar algo más allá de dichos correlatos.

Finalmente, allí donde la neurofenomenología se aproxima al ámbito ontológico (así, por ejemplo, al heredar la metafísica implícita en su asunción de la existencia de un problema duro y una brecha explicativa, y, particularmente, al enmarañarse en la discusión acerca de la reductibilidad) se muestra problemática y prematura, y muy probablemente sea lo primero el caso a causa de lo segundo. No obstante, como hemos destacado en sucesivas ocasiones, no dejaría éste de ser un problema compartido por todas y cada una de las propuestas que se aproximan desde este flanco al problema de la conciencia.

No quisiéramos cerrar estas tentativas apreciaciones críticas sin sumarnos al juicio de José Antonio Guerrero del Amo, que tras mostrar sus reservas y analizar algunas de las limitaciones del programa neurofenomenológico, pone de relieve que el mismo presenta un cariz "mucho más rico y creativo que los que encontramos habitualmente en las ciencias cognitivas" (Guerrero del Amo, 2012: 279) y, añadamos, asimismo más prometedor, bien que, exclusivamente, dentro del ámbito del estudio de la experiencia consciente *humana*.

# CONCLUSIONES

La primera y más importante conclusión que cabe extraer de nuestro recorrido por los *Consciousness Studies* es, parafraseando a Popper, la de la miseria del misterianismo: muestra consistentemente la historia que las más de las veces chocaron los profetas con la incomprensión –y a menudo no sólo con ella– de la práctica totalidad de su audiencia, carente de línea directa con el futuro o la trascendencia. Un profeta al revés habrá de correr una suerte pareja, quizá la contraria: si cuesta arriba solieron caminar los profetas del pasado, quizá cuesta abajo vengan haciéndolo estos antiprofetas coetáneos nuestros. Más allá de la sana chanza y la metáfora, la idea de que una explicación de la conciencia será por siempre algo inalcanzable resulta sólo dogmáticamente defendible. Analizamos pormenorizadamente su inconsistencia en el último capítulo de la primera parte. Añadamos aquí que no sabemos si *nuestras* actuales explicaciones de, por ejemplo, la génesis del universo podrían en algún caso ser tenidas por *la* explicación de ese proceso, pero no dejaremos de elaborarlas y reelaborarlas, avanzando eventualmente con ello hacia el ideal de esa explicación (en singular). Lo mismo, exactamente, sucede con cualquier otro objeto de estudio. Podemos empeñarnos cuanto queramos en que nuestras empíreas nociones predilectas, la de vida, pongamos por caso, pertenezcan a un indeterminado más allá de cualquier indagación posible, pero seguirá habiendo quienes, como siempre, se dediquen a partirlas en virtuales pedacitos para analizarlos, cotejarlos, integrarlos en diferentes contextos controlando sistemáticamente cada fuente de variación y reconstruyendo tentativamente sectores de la indefinida enormidad de partida. Es a eso a lo que se llama ciencia, y también filosofía. Es a eso, en definitiva, a lo que tradicionalmente se denominó actividad intelectual, una actividad consistente en el intento de avanzar hacia formas más agudas y profundas de comprensión. Podemos intentar dar en la diana y coadyuvar así a esa tarea, pero podemos también paralogizar y urdir sofismas destinados a persuadir de la inutilidad de jugar a los dardos –v. g., "sabemos

desde los eleáticos que el movimiento es imposible, luego los dardos no pueden moverse y, por tanto, nadie podrá acertar nunca en ninguna diana"–. Cada uno habrá de decidir por sí mismo cuál de ambas opciones se le presenta como más razonable. Desde luego, acertar en la diana es algo que sucede raras veces y así, como en los albores de los *Consciousness Studies* dijera Christof Koch, "no nos sorprendería que se probara la inadecuación de nuestros modelos, pero sólo formulándolos podremos al menos contribuir a clarificar las cuestiones que la siguiente generación de teorías tendrá que abordar" (Koch, 1993: 13τ). Poco antes Crick & Koch (1990: 264) habían argüido que ninguna teoría podrá explicar por sí misma, en solitario, todo lo que una ciencia de la conciencia debe explicar. Hemos defendido en esta tesis que esta intuición sigue siendo tan válida hoy como lo era hace veinticinco años. Sin embargo, nadie deja de buscar la felicidad cuando se le informa de que no podrá habérselas él solito con toda la del mundo: seguir desarrollando teorías es la única vía constructiva entre las abiertas, aunque acabe por hacerse manifiesta la incapacidad de las mismas para dar acabada cuenta del dominio de fenómenos que procuraban cubrir. Algunas serán descartadas, otras sólo podrán acometer labores más modestas que aquéllas que inicialmente pretendiera asignárselas, y "ten[dr]emos la impresión de que tratamos de ver el mundo entero por el ojo de una cerradura, o sencillamente de buscar la llave de ésta debajo del único farol que tenemos, [pero lo] cierto es que la propia búsqueda es, en sí misma, una tarea apasionante" (Rivière, 1987: 98), la única, por añadidura, capaz alumbrar formas más amplias y profundas de comprensión. Ningún argumento hegeliano parece poder hacer de esta humilde verdad una enrevesada mentira. De este modo, si, como sugiere la cita de Cioran con la que iniciamos esta tesis, la conciencia es la pesadilla de la naturaleza y, así, una pesadilla para el naturalismo, esa pesadilla, como todas, se desvanecerá cuando despertemos, como, por su parte, sugiere la de Goethe que la subsigue –aunque, ciertamente, también nosotros consideremos el de tornar la vista hacia el Sol un mal consejo.

Hemos defendido también que son muchas y muy diversas las tareas que al filósofo le cabe realizar en la investigación de vanguardia en ciencias cognitivas. Sostuvimos además que de entre esas tareas, la propiamente filosófica, la verdaderamente científica, la que cabría esperar de un enfoque naturalista consecuente, tendría prioritariamente que ver con la crítica, con el intento de hacer espacio para que la competencia entre las teorías en liza derive en un contexto de ampliadas posibilidades para la preponderancia de las más ponderadas entre las opciones disponibles. En lo atinente a este punto, pusimos de relieve el escaso provecho que cabe esperar extraer de la prosecución de la guerra

interparadigmática que ha venido entreteniendo a una sorprendente cantidad de teóricos en las dos últimas décadas e incidimos en que el oportunismo de una estrategia *catch-as-catch-can* puede resultar no sólo conveniente, sino quizá irrenunciable, necesario en vista no sólo del estado de nuestro conocimiento del objeto de los *Consciousness Studies*, sino asimismo del propio carácter de dicho objeto, muy probablemente el más vasto y multiforme de cuantos han de acomodar en su seno las ciencias de la mente, el cerebro y la conducta. Por otra parte, tras argumentar que el único problema de la conciencia es el explicativo, al que en último término se reducen todos y cada uno de los problemas de la conciencia que encontramos definidos y designados de distintos modos en la bibliografía –incluyendo, como subrayamos, el así llamado problema ontológico–, dividimos dicho problema en dos vertientes: la histórica y la funcional. Centramos en la segunda parte de la tesis nuestra atención en aquélla y vimos en la tercera que ninguno de los tres marcos teóricos globales más comentados en la bibliografía filosófica ofrece indicaciones particularmente precisas en esa dirección. Sea como fuere, en ambos casos, esto es, en lo tocante tanto al flanco histórico como al funcional, es aún evidentemente necesaria una gran cantidad de trabajo experimental y teórico, aunque quizá el primero sea el que requiera, en comparación, de más datos y esclarecimiento conceptual y teórico. No prestamos apenas atención al contraste entre teorías explicativas funcionales, sino que pusimos en ellas la mira sólo por su relación con la explicación histórica. Es en este contexto que cobra sentido el modo en que contrapusimos las teorías centradas antes bien en lo perceptivo y lo cognitivo a las teorías centradas en lo afectivo, destacando la necesidad de la integración dentro de la perspectiva explicativa histórica de las herramientas teóricas que en el marco de estas últimas vienen desarrollándose. No debe entenderse, por otra parte, que este énfasis en lo afectivo pretenda oponerse unilateralmente a la enorme cantidad de trabajo que en diferentes áreas de los *Consciousness Studies* ha venido realizándose desde el prisma cognitivista. La mente afectiva y la mente cognitiva no pueden ser irreconciliables porque una y otra no pasan de constituir meros descriptores de una realidad singular, aunque extraordinariamente plural, que difícilmente cabrá dividir limpiamente por esa línea imaginaria, porque de hecho no hay más que una sola mente, por más que frases como ésta parezcan desatender la innegable vaguedad de esta inmensa noción. La exuberante cantidad y calidad de los trastos que han venido siendo arrojados dentro de la misma hace que a no pocos se les dibuje una sonrisa burlona ante la singular locución «teoría de la mente» –siempre que no se emplee en el sentido técnico con que la acuñaran Premack y Woodruff a finales de los setenta–,

porque apenas nadie deja de admitir que "no hay un solo enfoque que parezca poder desvelar [por sí mismo y de forma aislada] el funcionamiento de la mente" (Johnson-Laird, 1988: 7 del original, 13 de la traducción). Obviar las aportaciones cognitivistas no se presenta pues como una alternativa razonable, entre otras cosas porque parece suficientemente probado que las formas superiores de la cognición influyen decisivamente en lo que experimentamos y, del mismo modo, en lo que pensamos que experimentamos. Lo cognitivo y lo afectivo pertenecen a diferentes niveles de análisis, lo cual no significa que esos niveles se encuentren desconectados. Con todo, no vislumbramos excesivo riesgo en la apuesta según la cual las teorías cognitivas de la conciencia seguirán gravitando antes sobre las explicaciones funcionales que sobre las históricas.

Puede que muchos consideren que las formas de conciencia a las que permiten acceder la neurofenomenología y la heterofenomenología son todo lo que una intachable biología de la conciencia debe explicar, pero las formas y dimensiones de la conciencia humana, tal y como pueden ser plasmadas en el lenguaje humano, son poco más o menos que la punta del iceberg del fenómeno que tal rama del saber humano ha de abordar. Sin embargo, sobra que insistamos en que tanto el fenómeno como la rama en cuestión son aquí referidos en singular pero concebidos en complicadísimo plural, como sobra que concedamos que es precisamente la forma humana de la experiencia consciente, mediada y afectada por el lenguaje y los procesos cognitivos superiores, un punto de partida privilegiado para el estudio de la biología de la conciencia. No obstante, afirmar que sólo de ella, sólo de la forma humana de la experiencia consciente tenemos noticias mínimamente dignas de crédito mientras que de cuanto reste más allá sólo nos cabe especular, se enfrenta a la más sencilla y eficaz entre las réplicas, pues precisamente eso son nuestras ciencias de hoy: las meras especulaciones de ayer.

**Referencias bibliográficas**


Adams, F. & Aizawa, K. (1994a) "Fodorian semantics", en S. Stich & T. Warfield (eds.), *Mental Representations*, Oxford: Basil Blackwell, pp. 223-242.

Adams, F. & Aizawa, K. (1994b) "'X' means X: Fodor/Warfield semantics", *Minds and Machines*, vol. 4, no. 2, pp. 215-231.

Albertazzi, L. (1996) "Edmund Husserl: 1859-1938", en L. Albertazzi, M. Libardi & R. Poli (eds.), *The School of Franz Brentano*, Dordrecht: Kluwer Academic Publishers, pp. 175-206.

Albright, T. D., Jessell, T. M., Kandel, E. R. & Posner, M. I. (2000) "Neural science: A century of progress and the mysteries that remain", *Cell*, vol. 100, *Neuron*, vol. 25 (suppl.), pp. 1-55.

Aleman, B. & Merker, B. (2014) "Consciousness without cortex: a hydranencephaly family survey", *Acta Paediatrica*, vol. 103, no. 10, pp. 1057-1065.

Allen, C. & Bekoff, M. (1997) *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. Cambridge, MA: MIT Press.

Alter, T. & Howell, R. J. (eds.) (2011) *Consciousness and The Mind-Body Problem: A Reader*. New York: Oxford University Press.

Ambrosio Flores, E. (2004) *Psicobiología de la drogadicción*. Madrid: Sanz y Torres.

Anderson, J. R. (1983) *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Angell, J. R. (1907) "The province of functional psychology", *Psychological Review*, no. 14, pp. 61-91. [Reimpreso en J. R. Shook (ed.), *The Chicago School of Functionalism, Vol. 3*, Bristol: Thoemmes Press, 2001, pp. 230-253].

Anscombe, G. E. M. (1965) "The intentionality of sensation: A grammatical feature", en R. J. Butler (ed.), *Analytical Philosophy: Second Series*, Oxford: Basil Blackwell, pp. 158-168. [Reimpreso en G. E. M. Anscombe, *The Collected*

*Philosophical Papers of G. E. M. Anscombe, Vol. 2. Metaphysics and the Philosophy of Mind*, Oxford, UK: Blackwell, 1981, pp. 3-20].

Århem, P., Liljenström, H. & Lindahl, B. I. B. (2002) "Evolution of consciousness. Report on the Agora workshop in Sigtuna, Sweden, on 11-13 August 2001", *Journal of Consciousness Studies*, vol. 9, no. 4, pp. 81–84.

Aristóteles (ca. 330 a. C.) *Acerca del alma*. Madrid: Gredos, 2003. [Trad. de T. Calvo Martínez].

Aristóteles (ca. 330 a. C.) *Segundos analíticos*, en *Tratados de lógica (Órganon) II*, Madrid: Gredos, 1995. [Trad. de M. Candel Sanmartín].

Aristóteles (ca. 330 a. C.) *Metafísica*. Madrid: Gredos, 2006. [Trad. de T. Calvo Martínez].

Armony, J. & Vuilleumier, P. (eds.) (2013) *The Cambridge Handbook of Human Affective Neuroscience*. New York: Cambridge University Press.

Armstrong, D. M. (1966) "The nature of mind", *Arts, Proceedings of the Sydney University Arts Association*, vol. 3, no. 1, pp. 37-48. [Reimpreso en D. M. Armstrong, *The Nature of Mind and Other Essays*, St. Lucia: University of Queensland Press, 1980, pp. 1-15].

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.

Armstrong, D. M. (1978a) "Naturalism, materialism, and first philosophy", *Philosophia*, vol. 8, nos. 2-3, pp. 261-276. [Reimpreso en D. M. Armstrong, *The Nature of Mind and Other Essays*, St. Lucia: University of Queensland Press, 1980, pp. 149-165].

Armstrong, D. M. (1978b) "What is consciousness?", *Proceedings of the Russellian Society*, vol. 3, pp. 65-76. [Reimpreso en D. M. Armstrong, *The Nature of Mind and Other Essays*, St. Lucia: University of Queensland Press, 1980, pp. 55-67].

Arrollo, E. (2016) *Ciencia y consciencia. La interacción entre mente y materia*. Barcelona: RBA.

Aserinsky , E. & Kleitman, N. (1953) "Regularly occurring periods of eye motility, and concomitant phenomena during sleep", *Science*, vol. 118, pp. 273-274.

Austin, J. L. (1956) "A plea for excuses", *Proceedings of the Aristotelian Society*, vol. 57, pp. 1-30. [Reimpreso en J. O. Urmson & G. J. Warnock (eds.), *Philosophical Papers*, New York: Oxford University Press, 1961, pp. 175-204; y, más

recientemente, en R. R. Ammerman (ed.), *Classics of Analytic Philosophy*, Indianapolis, IN: Hackett Publishing, 1990, pp. 379-398].

Baars, B. J. (1986) *The Cognitive Revolution in Psychology*. New York: Guilford Press.

Baars, B. J. (1988) *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.

Baars, B. J. (1991) "A curious coincidence? Consciousness as an object of scientific scrutiny fits our personal experience remarkably well", *Behavioral and Brain Sciences*, vol. 14, no.4, pp. 669-670.

Baars, B. J. (2005) "Global workspace theory of consciousness: toward a cognitive neuroscience of human experience?", *Progress in Brain Research*, vol. 150, pp. 45-53.

Baars, B. J. (2010) "Mind and brain", en B. Baars & N. Gage (eds.), *Cognition, Brain and Consciousness. Introduction to Cognitive Neuroscience* (2nd ed.), Amsterdam: Elsevier, pp. 3-32.

Bachelard, G. (1934) *Le Nouvel Esprit Scientifique*. Paris: Presses Universitaires de France. [Trad. de R. Sánchez, México, DF: Nueva imagen, 1981].

Baddeley, A. D. & Hitch, G. J. (1974) "Working memory", en G. A. Bower (ed.), *Recent Advances in Learning and Motivation, vol. 8*, New York: Academic Press, pp. 47-90.

Bailey, A. (2004) (ed.) *First Philosophy: Fundamental Problems and Readings in Philosophy. Vol III: God, Mind and Freedom*. Toronto: Broadview Press.

Bain, D. T. (2003) "Intentionalism and pain", *Philosophical Quarterly*, vol. 53, no. 213, pp. 502-523.

Balkin, T. J., Braun, A. R., Wesensten, N. J., Jeffries, K., Varga, M., Baldwin, P., et al. (2002) "The process of awakening: A PET study of regional brain activity patterns mediating the re-establishment of alertness and consciousness", *Brain*, vol. 125, no. 10, pp. 2308-2319.

Ballín, D. (1989) *El concepto de la conciencia*. México, DF: Fondo de Cultura Económica.

Balog, K. (1999) ''Conceivability, possibility, and the mind-body problem'', *Philosophical Review*, vol. 108, no. 4, pp. 497-528.

Balog, K. (2009) "Phenomenal concepts", en B. P. McLaughlin, A. Beckermann & S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind*, Oxford, UK: Oxford University Press, pp. 292-312.

Bartok, P. J. (2005) "Reading Brentano on the intentionality of the mental", en G. Forrai & G. Kampis (eds.), *Intentionality. Past and Future*, Amsterdam: Rodopi, pp. 15-24.

Bartra, R. (2006) *Antropología del cerebro. La conciencia y los sistemas simbólicos.* Valencia: Pre-textos.

Bayne, T. (2004) "Closing the gap? Some questions for neurophenomenology", *Phenomenology and the Cognitive Sciences*, vol. 3 no. 4, pp 349-364.

Bealer, G. (1994) "Mental properties", *The Journal of Philosophy*, vol. 91, no. 4, pp. 185-208.

Bechtel, W. (1998) "Representations and cognitive science: Assessing the dynamicist's challenge in cognitive science", *Cognitive Science*, vol. 22, no. 3, pp. 295-318.

Bechtel, W. (2006) *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge, UK: Cambridge University Press.

Bechtel, W. (2007) "Biological mechanisms: Organized to maintain autonomy", en F. Boogerd, F. J. Bruggeman, J-H. S. Hofmeyr & H. V. Westerhoff (eds.), *Systems Biology: Philosophical Foundations*, Amsterdam: Elsevier, pp. 269-302.

Bechtel, W. (2008) *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.

Bechtel, W. (2009) "Looking down, around, and up: Mechanistic explanation in psychology", *Philosophical Psychology*, vol. 22, no. 5, pp. 543-564.

Bechtel, W. & Abrahamsen, A. (2005) "Explanation: A mechanistic alternative", *Studies in History and Philosophy of the Biological and Biomedical Sciences*, vol. 36, no. 2, pp. 421-441.

Beck, F. & Eccles, J. C. (1992) "Quantum aspects of brain activity and the role of consciousness", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 23, pp. 11357-11361.

Beck, F. & Eccles, J. C. (2003) "Quantum processes in the brain: a scientific basis of consciousness", en N. Osaka (ed.), *Neural Basis of Consciousness*, Amsterdam & Philadelphia, PA: John Benjamin, pp. 141-166.

Beebee, H., Hitchcock, C. & Menzies, P. (eds.) (2009) *The Oxford Handbook of Causation*. Oxford, UK: Oxford University Press.

Benítez, A., Fernández, A. & del Arco, A. (2001) "Irreductibilidad de la conciencia en el naturalismo biológico de Searle", en J. M. Sagüillo, J. L. Falguera & C.

Martínez (eds.), *Actas del Congreso 'Teorías Formales y Teorías Empíricas'*, Santiago de Compostela: Universidad de Santiago de Compostela, pp. 255-270.

Bennett, M. R. (1997) *The Idea of Consciousness: Synapses and the Mind*. Amsterdam: Harwood.

Bennett, M. R. & Hacker, P. M. S. (2003) *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Berlin, I. (1939/1978) *Karl Marx. His Life and Environment*. Princeton, NJ: Princeton University Press, 2013. [Trad. de la tercera edición (de 1963) de R. Bixio, Madrid: Alianza, 1973].

Berlin, I. (1951) *The Hedgehog and the Fox: An Essay on Tolstoy's View of History*. London: Orion, 2009. [Trad. de C. Aguilar, Barcelona: Península, 2009].

Bermúdez, J. L. (2010/1014) *Cognitive Science. An Introduction to the Science of the Mind* (2nd ed.). Cambridge, UK: Cambridge University Press.

Bernabé Pajares, A. (2001) *De Tales a Demócrito. Fragmentos presocráticos* (2ª ed.). Madrid: Alianza.

Berry, J. W., Poortinga, E. H., Segall, M. H., Dasen, P. R. (eds.) (2002) *Cross-Cultural Psychology. Research and Applications* (2nd ed.). Cambridge, UK: Cambridge University Press.

Bickerton, D. (1995) *Language and Human Behaviour*. Seattle: University of Washington Press.

Bickle, J., Mandik, P. & Landreth, A. (2006) "The philosophy of neuroscience", en E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Spring 2006 Edition)*. URL: <http://plato.stanford.edu/archives/spr2006/entries/neuroscience/>.

Birbaumer, N., Murguialday, A. R., & Cohen, L. (2008) "Brain-computer interface in paralysis", *Current Opinion in Neurology*, vol. 21, no. 6, pp. 634-638.

Birbaumer, N., Weber, C., Neuper, C., Buch, E., Haapen, K., & Cohen, L. (2006) "Physiological regulation of thinking: Brain-computer interface (BCI) research", *Progress in Brain Research*, vol. 159, pp. 369-391.

Bjork, R. A. (1975) "Short-term storage: The ordered output of a central processor", en F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman & D. B. Pisoni (eds.), *Cognitive Theory Vol.1*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp.151-171.

Blackmon, J., Byrd, D., Cummins, R. C., Lee, A. & Roth, M. (2006) "Representation and unexploited content", en G. F. Macdonald & D. Papineau (eds.),

*Teleosemantics: New Philosophical Essays*, New York: Oxford University Press, pp. 195-207. [Reimpreso en R. C. Cummnins, *The World in the Head*, New York: Oxford University Press, 2010, pp. 120-134].

Blackmore, S. (2005) *Consciousness: A Very Short Introduction*. Oxford, UK: Oxford University Press.

Blackmore, S. (2006) *Conversations on Consciousness. What the Best Minds Think about the Brain, Free Will, and What It Means to Be Human*. Oxford, UK: Oxford University Press. [Trad. de F. Forn, Barcelona: Paidós, 2010].

Blake, R. & Logothetis, N. K. (2002) "Visual competition", *Nature Reviews. Neuroscience*, vol. 3, no.1, pp. 13-21.

Blanquet, P. R. (2011) "Advances in interdisciplinary researches to construct a theory of consciousness", *Journal of Behavioral and Brain Science*, vol. 1, pp. 242-261.

Block, N. (1978) "Troubles with functionalism", en C. W. Savage (ed.), *Perception and Cognition. Issues in the Foundations of Psychology. Minnesota Studies in the Philosophy of Science, vol. 9*, Minneapolis: University of Minneapolis Press, pp. 261-325.

Block, N. (1986) "Advertisement for a semantics for psychology", en P. French, T. Uehling & H. Wettstein (eds.), *Midwest Studies in Philosophy. Vol. 10. Studies in the Philosophy of Mind*, Minneapolis: University of Minnesota Press, pp. 615-678.

Block, N. (1990) "Inverted Earth", en J. E. Tomberlin (ed.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, Northridge: Ridgeview Publishing Company, pp. 53-79.

Block, N. (1993) "Consciousness ignored? Review of D. C. Dennett´s *Consciousness Explained*", *Journal of Philosophy*, vol. 90, no. 4, pp. 83-91.

Block, N. (1994a) "Consciousness", en S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Cambridge, MA: Blackwell, pp. 209-218.

Block, N. (1994b) "Qualia". en S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Cambridge, MA: Blackwell, pp. 514-520.

Block, N. (1994c) "What is Dennett's theory a theory of?", *Philosophical Topics*, vol. 22, nos. 1-2, pp. 23-40.

Block, N. (1995) "On a confusion about a function of consciousness", *Behavioral and Brain Sciences*, vol. 18, no. 2 pp. 227-247. [Reimpreso en N. Block, O.

Flanagan, & G. Güzeldere (eds.), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, 1997, pp. 375-415].

Block, N. (1996) "Mental paint and mental latex", en E. Villenueva, (ed.), *Philosophical Issues, 7: Perception*, Atascadero, CA: Ridgeview Publishing, pp. 19-49.

Block, N. (2001) "Paradox and cross purposes in recent work on consciousness", *Cognition*, vol. 79, no. 1, pp. 197-219.

Block, N. (2002) ''Some concepts of consciousness'', en D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, New York: Oxford University Press, pp. 206-218.

Block, N. (2003) "Consciousness, philosophical issues about", en L. Nadel (ed.), *Encyclopedia of Cognitive Science. Vol. 1*, London: Macmillan, pp. 760-770.

Block, N. (2007) "Consciousness, accessibility, and the mesh between psychology and neuroscience", *Behavioral and Brain Sciences*, vol. 30, nos. 5-6, pp. 481-499.

Bloom, P. (2004) *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*. New York: Basic Books.

Bodei, R. (1997) *La filosofia nel Novecento*. Roma: Donzelli. [Trad. de C. Caranci, Madrid: Alianza, 2001].

Boden, M. A. (ed.) (1990) *The Philosophy of Artificial Intelligence*. Oxford, UK: Oxford University Press. [Trad. de G. Feher de la Torre, México, DF: Fondo de Cultura Económica, 1994].

Boden, M. A. (2006) *Mind as Machine: A History of Cognitive Science.* New York: Oxford University Press.

Boole, G. (1854) *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton & Maberley. [Reimpreso en New York: Dover, 1961].

Brandl, J. (1996) "Intentionality", en L. Albertazzi, M. Libardi & R. Poli (eds.), *The School of Franz Brentano*, Dordrecht: Kluwer Academic Publishers, pp. 261-284.

Brandon, R. N. (1990) *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

Brentano, F. C. H. H. (1874) *Psychologie vom empirischen Standpunkte*, Leipzig: Duncker & Humblot [Ed. inglesa de O. Kraus (ed.), trad. de A. C. Rancurello, D. B. Terrell & L. L. McAlister, London: Routledge, 1995, por donde citamos].

Brewer, W. F. (1974) "There is no convincing evidence for operant or classical conditioning in adult humans", en W. B. Weimer & D. S. Palermo (eds.), *Cognition and the Symbolic Processes*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 1-42.

Brigandt, I. (2013) "Explanation in biology: Reduction, pluralism, and explanatory aims", *Science & Education*, vol. 22, no. 1, pp. 69-91.

Broad, C. D. (1925) *The Mind and Its Place in Nature*. London: Routledge & Kegan Paul. [Ed. actual: New York: Routledge, 2013].

Broadbent, D. E. (1958) *Perception and Communication*. New York: Pergamon Press.

Brook, A. (2005) "Making consciousness safe for neuroscience", en A. Brook & K. Akins (eds.), *Cognition and the Brain: The Philosophy and Neuroscience Movement*. New York: Cambridge University Press, pp. 397-422.

Brook, A. & Mandik, P. (2004) "The philosophy and neuroscience movement", *Analyse & Kritik*, vol. 26, pp. 382-397.

Brook, A. & Mandik, P. (2005) "Introduction", en A. Brook & K. Akins (eds.), *Cognition and the Brain: The Philosophy and Neuroscience Movement*. New York: Cambridge University Press, pp. 1-24.

Brooks, R. A. (1991) "Intelligence without representation", *Artificial Intelligence*, vol. 47, pp. 139-159.

Brown, R Glazebrook, J. F. & Baianu, I. C. (2007) "A conceptual construction of complexity levels theory in spacetime categorical ontology: Non-Abelian algebraic topology, many-valued logics and dynamic systems", *Axiomathes*, vol. 17, nos. 3-4, pp. 409-493.

Bruce Goldstein, E. (2010) *Sensation and Perception* (8th ed.). Belmont, CA: Wadsworth Publishing.

Budd, M. (1989) *Wittgenstein's Philosophy of Psychology*. London: Routledge.

Bunge, M. A. (1967/2004) *Scientific Research (2 vols.)*. New York: Springer. [Trad. de M. Sacristán, México, DF: Siglo XXI, 2004, por donde citamos; 3ª ed. corregida a partir de la 2ª, publicada en Barcelona: Ariel, 1983].

Bunge, M. A. (1980/1988) *The Mind-Body Problem: A Psychobiological Approach*. Oxford: Pergamon. [Versión castellana: Madrid: Tecnos, 2002].

Bunge, M. A. & Ardila, R. (1987) *Philosophy of Psychology*. New York: Springer. [Trad. de M. A. Galmarini, México, DF: Siglo XXI, 2002; ed. corregida a partir de la publicada en Barcelona: Ariel, 1988].

Burge, T. (1979) "Individualism and the mental", *Midwest Studies in Philosophy*, vol. 4, no. 1, pp. 73-122.

Buzsáki, G. (2006) *Rhythms of the Brain*. New York: Oxford University Press.

Byrne, A. (1994) "Behaviourism", en S. D. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Cambridge, MA: Blackwell, pp. 132-140.

Byrne, S. (2009) "Remarks on Ludwig Wittgenstein and behaviourism", en S. Nolan (ed.), *Maynooth Philosophical Papers, no. 5*, Maynooth: National University of Ireland Press, pp. 49-56.

Campbell, K. (1970) *Body and Mind*. London: Macmillan.

Canseco, J. (2007) "Redes neuronales y conexionismo en las neurociencias", *Metábasis*, vol. 2, no. 3, pp. 1-9.

Carmona, M. J. (2011) "Asexualidad, la vía más rápida para proliferar", *Investigación y Ciencia*, no. 414, pp. 14-15.

Carrier, M & Mittelstrass, J. (1995) *Mind, Brain, Behavior: The mind-Body Problem and the Philosophy of Psychology*. Berlin: Gruyter.

Carruthers, P. (1996) *Language, Thought and Consciousness: An Essay in Philosophical Psychology*. Cambridge, UK: Cambridge University Press.

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge, UK: Cambridge University Press.

Cartmill, M. (2000) "Animal consciousness: Some philosophical, methodological, and evolutionary problems", *Integrative and Comparative Biology (American Zoologist)*, vol. 40, no. 6, pp. 835-846.

Cavanna, A. E. (2008) "Seizures and consciousness", en S. C. Schachter, G. L. Holmes & D. G. A. Kasteleijn-Nolst Trenité (eds.), *Behavioral Aspects of Epilepsy: Principles and Practice*, New York: Demos, pp. 99-104.

Cavanna, A. E. & Nani, A. (2014) *Consciousness. Theories in Neuroscience and Philosophy of Mind*. Berlin: Springer.

Chalmers, A. F. (1976/2013) *What Is This Thing Called Science?* (4th ed.). St Lucia, Queensland: Queensland University Press. [Trad. española de la 2ª ed. inglesa de E. Pérez Sedeño y P. López Máñez, Madrid: Siglo XXI, 1984].

Chalmers, D. J. (1995) "Facing up to the problem of consciousness", *Journal of Consciousness Studies*, vol. 2, no. 3, pp. 200-219.

Chalmers, D. J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press [Trad. de J. A. Álvarez, Barcelona: Gedisa, 1999].

Chalmers, D. J. (2000), 'What is a neural correlate of consciousness?', en T. Metzinger (ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, Cambridge, MA: MIT Press, pp. 17-39.

Chalmers, D. J. (2002) "The puzzle of conscious experience", *Scientific American*, *Special*, vol. 12, no. 1, pp. 90-100.

Chalmers, D. (2004) "The representational character of experience", en B. Leiter (ed.), *The Future for Philosophy*, Oxford, UK: Oxford University Press, pp. 153-180.

Changeux, J-P. (1983) *L'Homme neuronal. Fayard Paris*. [Trad. de C. Janes, Madrid: Espasa Calpe, 1985].

Changeux, J-P. (2002) *L'Homme de vérité*. Paris: Odile Jacob. [Trad. de V. Aguirre, Mexico, DF: Fondo de Cultura Económica, 2004].

Changeux, J-P. (2008) *Du vrai, du beau, du bien: Une nouvelle approche neuronale*. Paris: Odile Jacob. [Trad. de J. Bucci, Buenos Aires: Katz, 2010].

Chemero, A. (2009) *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Chomsky, N. (1976) "Problems and mysteries in the study of human language", en A. Kasher (ed.), *Language in Focus: Foundations, Methods and Systems. Essays in Memory of Yehoshua Bar-Hillel*, Dordrecht: Reidel, pp. 281-357.

Chomsky, N. (1980) "Rules and representations", *Behavioral and Brain Sciences*, vol. 3, pp. 1-61.

Chomsky, N. (2000) *New Horizons in the Study of Language and Mind*. Cambridge, UK: Cambridge University Press.

Churchland, P. M. (1984/1987/2013) *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind* (3rd ed.). Cambridge, MA: MIT Press. [Trad. (de la segunda ed.) de M. N. Mizraji, Barcelona: Gedisa, 1992].

Churchland, P. M. (1988) "Folk psychology and the explanation of behavior," *Proceedings of the Aristotelian Society*, supl. vol. 62, pp. 209-221. [Reimpreso en P. M. Churchland, *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press, 1989, pp. 111-127, por donde citamos].

Churchland, P. M. (1989) *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*. Cambridge, MA: MIT Press.

Churchland, P. M. (1992) "Activation vectors versus propositional attitudes: How the brain represents reality", *Philosophy and Phenomenological Research*, vol. 52, no. 2, pp. 419-424. [Reimpreso en P. M. Churchland & P. S. Churchland, *On the contrary: critical essays, 1987-1997*, Cambridge, MA: Bradford/MIT Press, 1998, pp. 39-44]

Churchland, P. M. (1995) *The Engine of Reason, the Seat of the Soul. A Philosophical Journey into the Brain*. Cambridge, MA: The MIT Press.

Churchland, P. M. (1996) "The rediscovery of light", *The Journal of Philosophy*, vol. 93, no. 5, pp. 211-228.

Churchland, P. M. (1998) "Toward a cognitive neurobiology of the moral virtues", *Topoi*, vol. 17, no. 1, pp. 83-96. [Reimpreso en P. F. Martinez-Freire, (ed.), *Filosofia Actual de la Mente*, Madrid: Contrastes, supl. 6, 2001, pp. 259-289].

Churchland, P. M. (2005) "Functionalism at forty: A critical retrospective", *Journal of Philosophy*, vol. 102, no. 1, pp. 33-50. [Reimpreso en P. M. Churchland, *Neurophilosophy at Work*, New York: Cambridge University Press, 2007, pp. 18-36].

Churchland, P. M. & Churchland, P. S. (1990) "Could a machine think?", *Scientific American*, vol. 262, no. 1, pp: 32-37.

Churchland, P. M. & Churchland, P. S. (1997) "Recent work on consciousness: Philosophical, empirical and theoretical", *Seminars in Neurology*, vol. 17, pp. 101-108.

Churchland, P. M. & Churchland, P. S. (1998) *On the Contrary: Critical Essays, 1987-1997*. Cambridge, MA: Bradford/MIT Press.

Churchland, P. S. (1986) *Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P. S. (1998) "Brainshy: Nonneural theories of conscious experience", en S. R. Hameroff, A. W. Kaszniak & A. C. Scott (eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, Cambridge, MA: MIT Press, pp. 109-126.

Churchland, P. S. (2002) *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P. S. & Grush, R. (1999) "Computation and the brain", en R. A. Wilson & F. C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA: MIT Press, pp. 155-158.

Churchland, P. S. & Sejnowski, T. J. (1988) “Perspectives on cognitive neuroscience”, *Science*, vol. 242, no. 4879, pp. 741-745.

Churchland, P. S. & Sejnowski, T. J. (1992) *The Computational Brain*. Cambridge, MA: MIT Press.

Cíntora, A. (2005) *Los presupuestos irracionales de la racionalidad*. Barcelona: Anthropos.

Claparède, E. (1933) “La genèse de l'hypothèse”, *Archives de Psychologie*, vol. 24, pp. l-155.

Clark, A. (2001) *Mindware. An Introduction to the Philosophy of Cognitive Science*. New York: Oxford University Press.

Clark, A. (2002) “That special something: Dennett on the making of minds and selves”, en A. Brook & D. Ross (eds.), *Daniel Dennett*, Cambridge, UK: Cambridge University Press, pp. 187-205.

Clark, A. & Grush, R. (1999) “Towards a cognitive robotics”, *Adaptive Behavior*, vol. 7, no. 1, pp. 5-16.

Clark, A. & Toribio-Mateas, J. (1994) “Doing without representing?”, *Synthese*, vol. 101, no. 3, pp. 401-431.

Clark, G. & Riel-Salvatore, J. (2001) “Grave markers, middle and early upper paleolithic burials”, *Current Anthropology*, vol. 42, no. 4, pp. 481-490.

Clayton, N. S. & Dickinson, A. D. (1998) “Episodic-like memory during cache recovery by scrub jays”, *Nature*, vol. 395, no. 6699, pp. 272-274.

Clayton, N. S., Griffiths, D. P., & Dickinson, A. D. (2000) “Declarative and episodic-like memory in animals: Personal musings of a scrub jay”, en C. M. Heyes & L. Huber (eds.), *The Evolution of Cognition*, Cambridge, MA: MIT Press, pp. 273-288.

Cleeremans, A. (2005) “Computational correlates of consciousness”, en S. Laureys (ed.), *The Boundaries of Consciousness: Neurobiology and Neuropathology*, Amsterdam: Elsevier, pp. 81-98.

Cleeremans, A. (2008) “Consciousness: The radical plasticity thesis”, en R. Banerjee & B. K. Chakrabarti (eds.), *Models of Brain and Mind: Physical, Computational and Psychological Approaches. Progress in Brain Research. Vol. 168*, Amsterdam: Elsevier, pp. 19-33.

Colomina Almiñana, J. J. (2010) *Los problemas de las teorías representacionales de la conciencia*. La Laguna: Servicio de Publicaciones de la Universidad de La Laguna.

Copeland, J. (1993) *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.

Corballis, M. C. (2007) "The evolution of consciousness", en P. D. Zelazo, M. Moscovitch & E. Thompson (eds.), *The Cambridge Handbook of Consciousness*, New York: Cambridge University Press, pp. 571-595.

Corcoran, K. J. (2001) "The trouble with Searle's biological naturalism", *Erkenntnis*, vol. 55, no. 3, pp. 307-324.

Cortázar, J. F. (1962) *Historias de cronopios y de famas*. Buenos Aires: Minotauro. [Ed. actual, por la que citamos: Buenos Aires: Alfaguara, 1995].

Craik, K. J. W. (1943) *The Nature of Explanation*. Cambridge, UK: Cambridge University Press.

Crane, T. (1995/2003) *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representation* (2nd ed.). London: Routledge [Trad. de J. Almela, Mexico, DF: Fondo de Cultura Económica, 2008].

Crane, T. (2000) "The origins of qualia", en T. Crane & S. Patterson (eds.), *History of the Mind-Body Problem*, New York: Routledge, pp. 169-194.

Crane, T. (2003) "The intentional structure of consciousness", en Q. Smith & A. Jokic (eds.), *Consciousness: New Philosophical Perspectives*, Oxford, UK: Oxford University Press, pp. 33-56.

Crick, F. H. C. (1989) "Neural Edelmanism", *Trends in Neuroscience*, vol. 12, no. 7, pp. 240-248.

Crick, F. H. C. (1994) *The Astonishing Hypothesis: The Scientific Search For The Soul*. New York: Scribner. [Trad. de F. P. de la Cadena, Madrid: Debate, 1994].

Crick, F. H. C., & Koch, C. (1990) "Towards a neurobiological theory of consciousness", *Seminars in the Neurosciences*, vol. 2, pp. 263-275. [Reimpreso en N. Block, O. Flanagan, & G. Güzeldere (eds.), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, 1997, pp. 277-292].

Crick, F. H. C., & Koch, C. (1998) "Consciousness and neuroscience", *Cerebral Cortex*, vol. 8, no. 2, pp. 97-107.

Cummins, R. C. (1989) *Meaning and Mental Representation*. Cambridge, MA: MIT Press.

Cummins, R. C. & Cummins, D. D. (eds.) (2000) *Minds, Brains, and Computers: The Foundations of Cognitive Science : An Anthology*. London: Blackwell.

Damasio, A. R. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam. [Trad. de P. Jacomet, Santiago de Chile: Andrés Bello, 1996].

Damasio, A. R. (1996) "The somatic marker hypothesis and the possible functions of the prefrontal cortex". *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, vol. 351, no. 1346, pp. 1413-1420.

Damasio, A. R. (1998) "Investigating the biology of consciousness", *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences. The conscious Brain: Abnormal and Normal*, vol. 353, no. 1377, pp. 1879-1882.

Damasio, A. R. (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace & Co.

Damasio, A. R. (2004) "Existo, luego pienso", en E. Punset, *Cara a cara con la vida, la mente y el Universo. Conversaciones con los grandes científicos de nuestro tiempo*, Barcelona: Destino, pp. 163-170.

Damasio, A. R. (2010) *Self Comes to Mind: Constructing the Conscious Brain*. New York: Pantheon. [Trad. de F. Meler Orti, Barcelona: Destino, 2010].

Damasio, A. R. & Carvalho, G. B. (2013) "The nature of feelings: evolutionary and neurobiological origins", *Nature Reviews Neuroscience*, vol. 14, no. 2, pp. 143-152.

Dardis, A. (2008) *Mental Causation: The Mind-Body Problem*. New York: Columbia University Press.

Daros, W. R. (2005) "John Searle: ¿La conciencia es una emergencia del cerebro?", *In Itinere. Publicación de Estudios Interdisciplinarios*, no. 2, pp. 129-156.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray. [Ed. actual, por la que citamos: J. Carroll (ed.), Peterborough, Ontario: Bradview Press, 2003]. [Ed. española recomendada, con introducción de Francisco J. Ayala: Trad. de A. Zulueta, Madrid: Universidad Nacional Autónoma de México; Consejo Superior de Investigaciones Científicas; Academia Mexicana de Ciencias; Los libros de la Catarata, 2009].

Davidson, D. H. (1970) "Mental events" en L. Foster & J. W. Swanson (eds.), *Experience and Theory*, London: Duckworth, pp. 79-91 [Reimpreso en D. H.

Davidson, *Essays on Action and Events*, Oxford, UK: Oxford University Press, 1980, pp. 207-225].

Davidson, D. H. (1987) "Knowing one's own mind", *Proceedings and Addresses of the American Philosophical Association*, vol. 60, no. 3, pp. 441-458.

Dawkins, C. R. (2006) *The God Delusion*. London: Bantam. [Trad. de R. Hernández Weigand, Madrid: Espasa, 2007].

Dawkins, C. R. (2009) *The Greatest Show on Earth. The Evidence for Evolution*. New York: Free Press. [Trad. de J. Fabregat Carrascosa, Madrid: Espasa Calpe, 2009].

Dawson, M. R. W., Dupuis, B. & Wilson, M. (2010) *From Bricks to Brains: The Embodied Cognitive Science of LEGO Robots*. Edmonton: Athabasca University Press.

De Caro, M. (2009) "Mysterianism and Skepticism", *Iris. European Journal of Philosophy and Public Debate*, vol. 1, no. 2, pp. 449-458.

Deacon, T. W. (2012) *Incomplete Nature: How Mind Emerged from Matter*. New York: Norton.

Dehaene, S. & Changeux, J-P. (2004) "Neural mechanisms for access to consciousness", en M. Gazzaniga (ed.), *The Cognitive Neurosciences* (3rd ed.), Cambridge, MA: MIT Press, pp. 1145-1157.

Dehaene S., Changeux J-P., Naccache L., Sackur J. & Sergent, C. (2006) "Conscious, preconscious, and subliminal processing: A testable taxonomy", *Trends in Cognitive Sciences*, vol. 10, no. 5, pp. 204-211.

Dehaene, S. & Naccache, L. (2001) "Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework", *Cognition*, vol. 79, nos. 1-2, pp. 1-37.

Dennett, D. C. (1969/1986/2010) *Content and Consciousness*. New York: Routledge. [Trad. de J. M. Lebrón, Barcelona: Gedisa, 1996].

Dennett, D. C. (1975) "Why the law of effect will not go away", *Journal of the Theory of Social Behavior*, vol. 5, no. 2, pp. 169-187. [Reimpreso en D. C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press, 1978/1981, pp. 71-89].

Dennett, D. C. (1978) "Toward a cognitive theory of consciousness", *Minnesota Studies in the Philosophy of Science*, vol. 9, pp. 201-228. [Reimpreso en D. C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press, 1978/1981, pp. 149-173].

Dennett, D. C. (1983) "Intentional systems in cognitive ethology: The 'Panglossian Paradigm' defended", *The Behavioral and Brain Sciences*, vol. 6, no. 3, pp. 343-390. [Reimpreso en D. C. Dennett, *The Intentional Stance*. Cambridge, MA: MIT Press, 1987, pp. 237-268].

Dennett, D. C. (1987) *The Intentional Stance*. Cambridge, MA: MIT Press. [Trad. de D. Zadunaisky, Barcelona: Gedisa, 1991].

Dennett, D. C. (1988) "When philosophers encounter artificial intelligence", *Daedalus*, vol. 117, no. 1, pp. 283-295. [Reimpreso en D. C. Dennett, *Brainchildren: Essays on Designing Minds*, Cambridge, MA: MIT Press/Bradford Books, 1998, pp. 265-276].

Dennett, D. C. (1991a) *Consciousness Explained*. Boston: Little, Brown & Co. [Trad. de S. Balari Ravera, Barcelona: Paidós, 1995].

Dennett, D. C. (1991b) "The brain and its boundaries. Review of Colin McGinn's *The Problem of Consciousness: Essays Toward a Resolution*", *Times Literary Supplement*, vol. 10, no. 4.597.

Dennett, D. C. (1993a) "Review of Searle, *The Rediscovery of the Mind*", *Journal of Philosophy*, vol. 60, no. 4, pp. 193-205.

Dennett, D. C. (1993b) "Back from the drawing board", en B. Dahlbom (ed.), *Dennett and his Critics*, Oxford: Blackwell, pp. 203-235.

Dennett, D. C. (1994a) "Get real. Reply to my critics", *Philosophical Topics*, vol. 22, nos. 1-2, pp. 505-568.

Dennett, D. C. (1994b) "Instead of qualia", en A. Revonsuo & M. Kamppinen (eds.), *Consciousness in Philosophy and Cognitive Neuroscience*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 129-139.

Dennett, D. C. (1994c) "Self-portrait", en S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Cambridge, MA: Blackwell, pp. 236-244. [Reimpreso en D. C. Dennett, *Brainchildren: Essays on Designing Minds*, Cambridge, MA: MIT Press/Bradford Books, 1998, pp. 355-366].

Dennett, D. (1994d) "Tiptoeing past the covered wagons: A response to Carr", en *Emory Cognition Project, Report 28, 'Dennett and Carr Further Explained: an exchange'*, Department of Psychology, Emory University, Apr. 1994, pp. 1-18.

Dennett, D. C. (1996a). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books. [Trad. de F. Páez de la Cadena, Madrid: Debate, 2000].

Dennett, D. C. (1996b) "Qualia", en M. S. Gazzaniga (ed.), *Conversations in the Cognitive Neurosciences*, Cambridge, MA: MIT Press, pp. 175-193.

Dennett, D. C. (2001) "The zombie hunch: extinction of an intuition?", en A. O'Hear (ed.), *Philosophy at the New Millenium*, no. 48 (suppl.) del *Royal Institute of Philosophy*, Cambridge, UK: Cambridge University Press, pp. 27-43.

Dennett, D. C. (2003) *Freedom Evolves*. New York: Viking Press. [Trad. de R. Vilà Vernis, Barcelona Paidós, 2004].

Dennett, D. C. (2004) "No hay ningún responsabe", en E. Punset, *Cara a cara con la vida, la mente y el Universo. Conversaciones con los grandes científicos de nuestro tiempo*, Barcelona: Destino, pp. 173-182.

Dennett, D. C. (2005) *Sweet Dreams*: *Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press. [Trad. de J. Barba & S. Jawerba, Buenos Aires/Madrid: Katz, 2006].

Dennett, D. C. (2006) *Breaking the Spell: Religion as a Natural Phenomenon*. New York: Viking. [Trad. de F. De Brigard, Buenos Aires/Madrid: Katz, 2007].

Dennett, D. C. (2007) "The science studio with Daniel Dennett", entrevista de R. Bingham. URL: <http://thesciencenetwork.org/media/videos/29/Transcript.pdf>.

Dennett, D. C. & Haugeland, J. (1987) "Intentionality", en R. L. Gregory (ed.), *The Oxford Companion to the Mind*, Oxford, UK: Oxford University Press, pp. 383-386. [Trad. de I. Cifuentes de Castro et al., revisión técnica de J. Cordero, Madrid: Alianza, 1995].

Denton, D. A. (1993) *The Pinnacle of Life. Consciousness and Self-awareness in Humans and Animals*. Crows Nest, New South Wales: Allen and Unwin.

Denton, D. A. (2005) *The Primordial Emotions: The Dawning of Consciousness*. Oxford, UK: Oxford University Press. [Trad. de J. Ros, Barcelona: Paidós, 2009].

Denton, D. A., McKinley, M. J., Farrell, M., & Egan, G. F. (2009) "The role of primordial emotions in the evolutionary origin of consciousness", *Consciousness and Cognition*, vol. 18, no. 2, 500-514.

Depraz, N., Varela, F. J. & Vermersch, P. (2003) *On Becoming Aware: A Pragmatics of Experiencing*. Amsterdam: John Benjamins Publishing Co.

Descartes, R. (1637) *La Dioptrique*, en F. Alquié (ed.), *Descartes, Oeuvres Philosophiques*, Tomo I, Paris: Garnier, 2001 (traducción francesa del original en latín aprobada por Descartes y publicada por vez primera en 1647).

Descartes, R. (1641) *Les Méditations, les Objections et les Réponses*, en F. Alquié (ed.), *Descartes, Oeuvres Philosophiques*, Tomo II, Paris: Garnier, 1967 (traducción francesa del original en latín aprobada por Descartes y publicada por vez primera en 1647). [Trad., prólogo y notas de V. Peña, Oviedo: KRK, 2005].

Descartes, R. (1644) *Les Principes de la Philosophie*, en F. Alquié (ed.), *Descartes, Oeuvres Philosophiques*, Tomo III, Paris: Garnier, 1971 (traducción francesa del original en latín aprobada por Descartes y publicada por vez primera en 1647). [Trad. de G. Quintás Alonso, Madrid: Alianza, 1995].

Díaz, J. L. (2007) *La conciencia viviente*, México, DF: Fondo de Cultura Económica.

Díaz, J. L. (2008) "La conciencia y el cerebro: a propósito de 'La Flama Misteriosa'", *Salud mental*, vol. 31, no. 3, pp. 239-246.

Díaz, J. L. & Velázquez, D. N. (2000) "La discriminación del efecto de las drogas y la conciencia animal", Salud Mental, vol. 23, no. 2, pp. 1-7.

Dickinson, A. & Balleine, B. W. (2000) "Causal cognition and goal-directed action", en C. M. Heyes & L. Huber (eds.), *The Evolution of Cognition*, Cambridge, MA: MIT Press, pp. 185-204.

Diéguez Lucena, A. (2013) "De nuevo, la mente como excepción. Algunos comentarios críticos acerca del antinaturalismo de Thomas Nagel", *Ludus Vitalis*, vol. 21, no. 39, pp. 343-354.

Diéguez Lucena, A. (2014) "Delimitación y defensa del naturalismo metodológico (en la ciencia y la filosofía)", en R. Gutiérrez Lombardo & J. Sanmartín Esplugues (eds.), *La filosofía desde la ciencia*, Mexico, DF: Centro de Estudios Filosóficos, Políticos y Sociales Vicente Lombardo Toledano, pp. 21-49.

Dobzhansky, T. G. (1973) "Nothing in biology makes sense except in the light of evolution", *The American Biology Teacher*, vol. 35, no. 3, pp. 125-129.

Domjan, M. P. (2015) *The Principles of Learning and Behavior* (7th ed.). Stamford, CT: Cengage. [Trad. de la 5ª ed. inglesa de H. Matute, R. Pellón, G. Muriel Good & M. A. Vallido, Madrid: Paraninfo, 2009].

Donnerer, J. & Lembeck, F. (2006) *Chemical Languages of the Nervous System. History of Scientists and Substances*. Basel: Karger.

Dretske, F. I. (1981) *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. I. (1983) "Précis of *Knowledge and the Flow of the Information*" The Behavioral and Brain Sciences, vol. 6, no. 1, pp. 55-63.

Dretske, F. I. (1995) *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Dretske, F. I (1996) "Phenomenal externalism, or If meanings ain't in the head, where are qualia?", en E. Villanueva (ed.), *Philosophical Issues, 7: Perception*, Atascadero, CA: Ridgeview Publishing, pp. 143-158.

Dretske, F. I. (2003) "How do you know you are not a zombie?", in B. Gertler (ed.), *Privileged Access and First Person Authority*, Burlington, VT: Ashgate Publishing, pp. 1-13.

Dreyfus, H. L. (1979) *What Computers Can't Do* (2nd ed.). New York: Harper and Row.

Dreyfus, H. L. & Kelly, S. D. (2007) "Heterophenomenology: Heavy-handed sleight-of-hand", *Phenomenology and the Cognitive Sciences*, vol. 6, nos. 1-2, pp. 45-55.

Drozdov, A. L. (2003) "Multikingdom system of living organisms", *Russian Journal of Nematology*, vol. 11, no. 2, pp. 127-132.

Earl, B. (2014) "The biological function of consciousness", *Frontiers in Psychology*, vol. 5, art. 697. DOI: <10.3389/fpsyg.2014.00697>.

Eccles, J. C. (1953) *The Neurophysiological Basis of Mind. The Principles of Neurophysiology*. Oxford, UK: Clarendon Press.

Eccles, J. C. (1974) "Cerebral activity and consciousness", en F. J. Ayala & T. G. Dobzhansky (eds.), *Studies in the Philosophy of Biology. Reduction and Related Problems*, Berkeley: University of California Press, pp. 87-107.

Eccles, J. C. (1980) *The Human Psyche*. New York: Springer.

Eccles, J. C. (1981) "The modular operation of the cerebral neocortex considered as the material basis of mental events", *Neuroscience*, vol. 6, no. 10, pp. 1839-1856.

Eccles, J. C. (1990) "A unitary hypothesis of mind-brain interaction in the cerebral cortex", *Proceedings of the Royal Society, Series B: Biological Sciences*, vol. 240, no. 1299, pp. 433-451.

Edelman, G. M. (1978) "Group selection and phasic reentrant signaling: A theory of higher brain function", en F. O. Schmitt & F. G. Worden (eds.) *The Neurosciences, IV Study Program*, Cambridge, MA: MIT Press, pp. 1115-1139. [Reimpreso en G. M. Edelman & V. B. Mountcastle, *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function*, Cambridge, MA: MIT Press, 1978/1982 pp. 51-100, por donde citamos].

Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Edelman, G. M. (1988) *Topobiology: An Introduction to Molecular Embriology*. New York: Basic Books.

Edelman, G. M. (1989) *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.

Edelman, G. M. (1992) *Bright Air, Brilliant Fire. On the Matter of the Mind*. New York: Basic Books.

Edelman, G. M. & Tononi, G. (2000) *A Universe of Consciousness: How Matter Becomes Imagination*. New York: Basic Books. [Trad. de J. L. Riera, Barcelona: Crítica, 2002].

Edelman, G. M. (2003) "Naturalizing consciousness: A theoretical framework", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 9, pp. 5520-5524.

Edelman, G. M. (2004). *Wider than the Sky: The Phenomenal Gift of Consciousness*. New Haven, CON: Yale University Press.

Eichenbaum, H. (2002/1012) *The Cognitive Neuroscience of Memory* (2nd ed.). Oxford, UK: Oxford University Press.

Einstein, A. (1936) "Physics and Reality", *Journal of the Franklin Institute*, vol. 221, no. 3, pp. 349-382. [Reimpreso en A. Einstein, *Ideas and Opinions*, New York: Crown Publishers, 1954, pp. 290-322 (traducido del alemán por S. Bargmann), por donde citamos. Trad. española de J. M. Álvarez, Barcelona: Antoni Bosch, 2011, pp. 289-322].

Eldredge, N. (2005) *Darwin: Discovering the Tree of Life*. New York: W. W. Norton & Co. [Trad. de J. Barba & S. Jawerbaum, Madrid: Katz, 2009].

Engel, A. K. & Singer, W. (2001) "Temporal binding and the neural correlates of sensory awareness", *Trends in Cognitive Sciences*, vol. 5, no. 1, pp. 16-25.

Engel, A. K., Fries, P. & Singer, W. (2001) "Dynamic predictions: oscillations and synchrony in top-down processing", *Nature Reviews Neuroscience*, vol. 2, no. 10, pp. 704-716.

Enríquez de Valenzuela, P. (2014a) "Conciencia", en P. Enríquez de Valenzuela (coord.), *Neurociencia cognitiva*, Madrid: Sanz y Torres, pp. 289-310.

Enríquez de Valenzuela, P. (2014b) "Lateralización hemisférica", en P. Enríquez de Valenzuela (coord.), *Neurociencia cognitiva*, Madrid: Sanz y Torres, pp. 249-266.

Enríquez de Valenzuela, P. (2014c) "Emoción", en P. Enríquez de Valenzuela (coord.), *Neurociencia cognitiva*, Madrid: Sanz y Torres, pp. 155-175.

Estany, A. (2005) "Bases teóricas de la explicación científica en la psicología", en A. Estany (ed.), *Filosofía de las ciencias naturales, sociales y matemáticas*, Madrid: Trota, pp. 261-292.

Eysenck, M. W. (2000) *Psychology: A Student's Handbook*. East Sussex: Taylor & Francis.

Faigenbaum, G. (2003) *Conversations with John Searle*. Montevideo: Libros en Red.

Feigl, H. (1958) "The 'mental' and the 'physical'", en H. Feigl, M. Scriven & G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science, Volume II*, Minneapolis: University of Minnesota Press, pp 370-497.

Feigl, H. (1981) *Inquiries and Provocations: Selected Writings, 1927–1974*, R. S. Cohen (ed.), London: Reidel.

Fernández-Guardiola, A. (1979) *La Conciencia: el problema mente-cerebro*. México, DF: Trillas.

Feynman, R. P. (1963) *Six Easy Pieces: Essentials of Physics Explained by its most Brilliant Teacher*. New York: Basic Books. [Trad. de J. García Sanz, Barcelona: Crítica, Grijalbo Mondadori, 1998].

Fields, D. R. (2006) "Beyond the neuron doctrine", *Scientific American Mind*, vol. 17, no. 3, pp. 21-27.

Finger, S. (1994) *Origins of Neurosciences: A History of Explorations into Brain Function*. New York: Oxford University Press.

Flanagan, O. J. (1984/1991) *The Science of the Mind* (2nd ed.). Cambridge, MA: MIT Press.

Flanagan, O. J. (1992) *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Fodor, J. A. (1974) "Special sciences (or: The disunity of science as a working hypothesis)", *Synthese*, vol. 28, no. 2, pp. 97-115.

Fodor, J. A. (1975) *The Language of Though*. New York: Crowell. [Trad. de J. Fernández, Madrid: Alianza, 1985].

Fodor, J. A. (1983) *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press. [Trad. de J. M. Igoa, Madrid: Morata, 1986].

Fodor, J. A. (1984) "Semantics, Wisconsin style", *Synthese*, vol. 59, no. 3, pp. 231-250. [Reimpreso en J. A. Fodor, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, 1990, pp. 30-49].

Fodor, J. A (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1989) "Making mind matter more", *Philosophical Topics*, vol. 17, no. 1, pp. 59–79. [Reimpreso en J. A. Fodor, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, 1990, pp. 137-160].

Fodor, J. A. (1990a) *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990b) "Information and representation", en P. Hanson (ed.), *Information, Language, and Cognition*, Vancouver: University of British Columbia Press, pp. 175-190.

Fodor, J. A. (1990c) "Psychosemantics or where do truth conditions come from?", en W. Lycan (ed.), *Mind and Cognition*, Oxford: Basil Blackwell, pp. 312-337.

Fodor, J. A. (1994) *The Elm and the Expert*. Cambridge, MA: MIT Press. [Trad. de M. A. Galmarini, Barcelona: Paidós, 1997].

Fodor, J. A. (2004) "You can't argue with a novel", *London Review of Books*, vol. 26, no. 5, pp. 30-31.

Fodor, J. A. & McLaughlin, B. P. (1990) "Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work", *Cognition*, vol. 35, no. 2, pp. 183-204.

Fodor, J. A. & Pylyshyn, Z. (1988) "Connectionism and cognitive architecture: A critical analysis", *Cognition*, vol. 28, nos. 1-2, pp. 3-71.

Fogelin, R. J. (1985) "The logic of deep disagreements", *Informal Logic,* vol. 7, no.1, pp. 1-8.

Foster, J. (1989) "A defense of dualism". J. Smythies & J. Beloff (eds.) *The Case for Dualism*. Charlottesville, VA: University of Virginia Press.

Foster J. (1991) *The Immaterial Self: A Defence of the Cartesian Dualist Conception of Mind.* London: Routledge.

Freeman, W. J. (1999) "Consciousness, intentionality, and causality", *Journal of Consciousness Studies*, vol. 6, no. 11-12, pp. 143–172.

Freeman, W. J. (2000a) "Emotion is essential to all intentional behaviors", en M. D. Lewis & I. Granic (eds.), *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*, Cambridge, UK: Cambridge University Press, pp. 209-235.

Freeman, W. J. (2000b) *Neurodynamics: An Exploration in Mesoscopic Brain Dynamics*. Berlin: Springer.

Freeman, W. J. & Skarda, C. A. (1991) "Mind/brain science: Neuroscience on philosophy of mind", en E. Lepore & R. Van Gulick (eds.), *John Searle and his Critics*, Oxford: Blackwell, pp. 115-127.

Frege, F. L. G. (1892) "On Sense and Reference", en P. Geach & M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 1980, pp. 56-78.

Freud, S. S. (1910) "Die psychogene Sehstörung in psychoanalytischer Auffassung", *Ärztliche Fortbildung* (suplemento de *Ärztlichen Standeszeitung*), vol. 9, no. 9, pp. 42-44. [Traducción de L. López-Ballesteros y de Torres, en S. S. Freud, *Obras Completas*, Madrid: Biblioteca Nueva, 1972, pp.1631-1635].

Frith, C. & Rees, G. (2007) "A brief history of the scientific approach to the study of consciousness", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 9-22.

García-Carpintero, M. (1995) "El funcionalismo", en F. Broncano (ed.), *La mente humana*, Madrid: Trotta, pp. 43-76.

García-Carpintero, M. (1996) *Las palabras, las ideas y las cosas. Una presentación de la filosofía del lenguaje*. Barcelona: Ariel.

García García, E. (2001) *Mente y cerebro*. Madrid: Síntesis.

García Suárez, A. (1995) "Qualia: propiedades fenoménicas", en F. Broncano (ed.), *La mente humana*, Madrid: Trotta, pp. 353-383.

García Valero, H. A. (2003) "La filosofía de la mente de John Searle", *Lógoi. Revista de Filosofía*, no. 6, pp. 41-82.

García Valero, H. A. (2006) "Reseña. Searle, John: 'El misterio de la conciencia'", *Lógoi. Revista de Filosofía*, no. 9, pp. 121-131.

Gardner, H. E. (1985) *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books. [Ed. de 1987 que incluye el epílogo "Cognitive science after 1984"]. [Trad. de L. Wolfson, Barcelona: Paidós, 1987].

Garrett, B. J. (1995) "Non-reductionism and John Searle's 'The Rediscovery of the Mind'", *Philosophy and Phenomenological Research*, vol. 55, no. 1, pp. 209-215.

Gasser, G. (ed.) (2007) *How Successful is Naturalism?* Frankfurt am Main: Publications of the Austrian Ludwig Wittgenstein Society. Ontos Verlag.

Gazzaniga, M. S. (1985) *The Social Brain: Discovering the Networks of the Mind*. New York: Basic Books. [Trad. de C. Frade Blas, Madrid: Alianza, 1993].

Gazzaniga, M. S. (1988) *Mind Matters: How Mind and Brain Interact to Create Our Conscious Lives*. Boston: Houghton Mifflin. [Trad. de D. Vidoni, Barcelona: Herder, 1998].

Gazzaniga, M. S. (1998) *The Mind's Past*. Berkeley & Los Angeles, CA: University of California Press. [Trad. de P. Jacomet, Santiago de Chile & Barcelona: Andrés Bello].

Gazzaniga, M. S. (ed.) (2000) *The New Cognitive Neurosciences* (2nd ed.). Cambridge, MA: MIT Press.

Gazzaniga, M. S. (2005) "Forty five years of split-brain research and still going strong", *Nature Reviews Neuroscience*, vol. 6, no. 8, pp. 653-659.

Gazzaniga, M. S. (2008) *Human: The Science behind what Makes us Unique*. New York: Harper Collins. [Trad. de F. Forn, Barcelona: Paidós].

Gazzaniga, M. S., LeDoux, J. E., & Wilson, D. H. (1977) "Language, praxis, and the right hemisphere: clues to some mechanisms of consciousness", *Neurology*, vol. 27, no.12, pp. 1144-1147.

Geisz, S. F. (2009) "Turning representation inside out: An adverbial approach to the metaphysics of language and mind", *The Philosophical Forum*, vol. 40, no. 4, pp 437-471.

Gentner, D. (2010) "Psychology in cognitive science: 1978–2038", *Topics in Cognitive Science*, vol. 2, no. 3, pp. 328-344.

Gertler, B. (2001) "The relationship between phenomenality and intentionality: Comments on Siewert's *The Significance of Consciousness*", *Psyche*, vol. 7, no. 17. URL: <http://www.theassc.org/files/assc/2508.pdf>.

Gibb, S. C., Lowe, E. J. & Ingthorsson, R. D. (eds.) (2013) *Mental Causation and Ontology*. Oxford, UK: Oxford University Press.

Gibson, M., (1996) "Asymmetric dependencies, ideal conditions, and meaning", *Philosophical Psychology*, vol. 9, no. 2, pp. 235-259.

Glock, H-J. (2008) *What is Analitic Philosophy?* Cambridge, UK: Cambridge University Press.

Godfroid, I. O. (2003) "Psychiagenia: A gauge theory for the mind-brain problem", *NeuroQuantology*, vol. 1, no. 2, pp. 189-199.

Gold, I. & Stoljar, D. (1999) "A neuron doctrine in the philosophy of neuroscience", *Behavioral and Brain Sciences*, vol. 22, no. 5, pp. 809-830.

Gomila Benejam, A. (2007) "Los laberintos de la filosofía de la mente: un mapa de la situación", en D. Pérez Chico & M. Barroso (eds.), *Pluralidad de la filosofía analítica*, Madrid: Plaza y Valdés, pp. 189-216.

González, J. (2008) "Filosofía y ciencias cognitivas", *Inventio*, no. 8, pp. 59-67.

González, J. (2009) "El papel del filósofo frente a las ciencias cognitivas", *Inventio*, no. 9, pp. 67-71.

González, M. (1989) *Introducción al pensamiento filosófico. Filosofía y modernidad*. Madrid: Tecnos.

González Álvarez, J. (2010) *Breve historia del cerebro*. Barcelona: Crítica.

González-Castán, O. L. (1999) "The connection principle and the classificatory scheme of reality", *Teorema*, vol. 18, no. 1, pp. 85-98.

González Labra, M. J. (2011) *Introducción a la psicología del pensamiento* (7ª ed.). Madrid: Trotta.

González Recio, J. L. (2007) "Aire, calor y sangre o la vida inventada desde el Mediterráneo", en J. L. González Recio (ed.), *Átomos, almas y estrellas: estudios sobre la ciencia griega*, Madrid: Plaza y Valdés, pp. 147-200.

Gould, J. L. (1990) "Honey bee cognition", Cognition, vol. 37, nos. 1-2, pp. 83-103.

Gould, J. L. & Gould, C. G. (1988/1995) *The Honey Bee* (2nd ed.). New York: W. H. Freeman.

Gould, S. J. (1985) *The Flamingo's Smile: Reflections in Natural History*. New York: W. W. Norton & Co. [Trad. de A, Resines, Barcelona: Crítica, Drakontos, 2008].

Gray, E. (1959) "Axosomatic and axodendritic synapses in the cerebral cortex: An electron microscope study", *Journal of Anatomy*, vol. 93, pp. 420-433.

Gray, J. A. (2004) *Consciousness: Creeping Up on the Hard Problem*. Oxford, UK: Oxford University Press.

Grice, H. P. (1975) "Logic and conversation", en P. Cole & J. L. Morgan (eds.), *Syntax and Semantics, Volume 3: Speech Acts*, New York: Academic Press, pp. 41-58.

Gregory, R. L., (ed.) (1987) *The Oxford Companion to the Mind*. Oxford, UK: Oxford University Press. [Trad. de I. Cifuentes de Castro et al., revisión técnica de J. Cordero, Madrid: Alianza, 1995].

Gregory, R. L. (2001) "Oliver Louis Zangwill. 29 October 1913 – 12 October 1987", *Biographical Memoirs of Fellows of the Royal Society*, vol. 47, pp. 515-524.

Gross, R. (2012) *Being Human: Psychological and Philosophical Perspectives*. New York: Routledge.

Gu, F., Meng, X., Shen, E. & Cai, Z. (2003) "Can we measure consciousness with EEG complexities?", *International Journal of Bifurcation and Chaos*, vol. 13, no. 3, pp. 733-742.

Guarino, J. I. (2010) "La relación causal entre la mente y el cuerpo. Searle y el «naturalismo biológico»", *Eikasia. Revista de Filosofía*, año V, no. 32, pp. 33-41.

Guerrero del Amo, J. A. (2000) "La naturalización de la epistemología en Hume", *Revista de Filosofía*, vol. 12, no. 23, pp. 61-84.

Guerrero del Amo, J. A. (2001) "Problemas epistemológicos subyacentes a la teoría de la mente de Searle", *Logos: Anales del seminario de metafísica*, vol. 34, no. 3, pp. 297-316.

Guerrero del Amo, J. A. (2012) ¿Es la neurofenomenología la solución al problema de la conciencia?, *Thémata. Revista de Filosofía*, no. 46, pp. 271-279.

Guillot, P. (2010) "Cartesian echoes in the philosophy of mind: The case of John Searle", en J. Reynolds, J. Chase, J. Williams & E. Mares, E. (eds.), *Postanalytic and Metacontinental: Crossing Philosophical Divides*, London: Continuum, pp. 107-124.

Gunson, D. (1998) *Michael Dummett and the Theory of Meaning*. Avebury Series in Philosophy. Aldershot: Ashgate.

Güzeldere, G. (1995) "Is consciousness the perception of what passes in one's own mind?", en T. Metzinger (ed.), *Conscious Experience*, Paderborn: Ferdinand Schoningh, pp. 335-358. [Reimpreso en N. Block, O. Flanagan, & G. Güzeldere (eds.), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, 1997, pp. 789-807].

Güzeldere, G. (1997) "The many faces of consciousness", en N. Block, O. Flanagan, & G. Güzeldere (eds.), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, pp. 1-67.

Haack, S. (2009). *Evidence and Inquiry: A Pragmatist Reconstruction of Epistemology* (2nd ed.). Amherst, New York: Prometheus Books.

Hacker, P. M. S. (2002) "Is there anything it is like to be a bat?", *Philosophy*, vol. 77, no. 2, pp. 157-174.

Hales, C. (2010) "The scientific evidence of qualia meets the qualia that are scientific evidence", *Psyche*, vol. 16, no. 1, pp. 24-29.

Hameroff, S. R. (2014) "Consciousness, microtubules, & 'Orch OR'. A 'Space-time Odyssey'", *Journal of Consciousness Studies*, vol. 21, nos. 3-4, pp. 126–153.

Hameroff, S. R. & Penrose, R. (2014) "Consciousness in the universe: A review of the 'Orch OR' theory", *Physics of Life Reviews*, vol. 11, no. 1, pp. 39-78.

Hanson, N. R. (1958) *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge, UK: Cambridge University Press [Trad. de E. García Camarero, Madrid: Alianza, 1977].

Harley, T. A. (2014) *The Psychology of Language: From Data to Theory* (4th ed.). Hove: Psychology Press. [Trad. de la tercera ed. inglesa de Y. Moreno López, Madrid: McGraw-Hill, 2009].

Harman, G. (1990) "The intrinsic quality of experience", en J. E. Tomberlin (ed.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, Atascadero, CA: Ridgeview Publishing, pp. 31-52. [Reimpreso en N. Block, O. Flanagan & G. Guzeldere (eds.) (1997), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, pp. 663–676].

Harris, E. E. (2006) *Reflections on the Problem of Consciousness*. Dordrecht, The Netherlands: Springer.

Haselager, P., de Groot, A., & van Rappard, H. (2003) "Representationalism vs. anti-representationalism: A debate for the sake of appearance", *Philosophical Psychology*, vol. 16, no. 1, pp. 5-23.

Hasker, W. (1999) *The Emergent Self*. Ithaca, New York: Cornell University Press.

Haugeland J. (ed.) (1997) *Mind Design II*. Cambridge, MA: MIT Press.

Hawking, S. W. & Mlodinow, L. (2010) *The Grand Design*. New York: Bantam Books. [Trad. de D. Jou i Mirabent, Barcelona: Crítica, 2010].

Hayek, F. A. (1952) *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*. Chicago: University of Chicago Press.

Hebb, D. O. (1949) *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley & Sons. [Ed. actual: Mahwah, NJ: Lawrence Erlbaum Associates, 2003].

Hebb, D. O. (1960) "The american revolution", *American Psychologist*, vol. 15, no. 12, pp. 735-745.

Hebb, D. O. (1974) "What psychology is about", *American Psychologist*, vol. 29, no. 2, pp. 71-79.

Heider, E. R. (1971) "'Focal' color areas and the development of color names", *Developmental Psychology*, vol. 4, no. 3, pp. 447-455.

Heider, E. R. (1972) "Universals in color naming and memory", *Journal of Experimental Psychology*, vol. 93, no. 1, pp. 10-20.

Heider, E. R. & Olivier, D. C. (1972) "The structure of color space in naming and memory for two languages", *Cognitive Psychology*, vol. 3, no. 2, pp. 337-354.

Heil, J. (2004) *Philosophy of Mind: A Contemporary Introduction* (2nd ed.). New York: Routledge.

Hendrichs, H. (1999) "Accessing animal minds: Epistemological and empirical problems", *Evolution and Cognition*, vol. 5, no. 2, pp. 98-104.

Hergenhahn, B. R. & Henley, T. B. (2014) *An Introduction to the History of Psychology* (7th ed.). Belmont, CA: Wadsworth.

Hermoso, J. & Chacón, P. (2000) "Sobre el carácter irreductible de la intencionalidad: la ontología del inconsciente y los dos conceptos de trasfondo en Searle", en P. Chacón Fuertes & M. Rodríguez González (eds.), *Pensando la mente. Perspectivas en filosofía y psicología*, Madrid: Biblioteca Nueva, pp. 167-194.

Hernández Guijo, J. M. (2008) "La naturaleza química de la transmisión sináptica. Un largo camino hacia el neurotransmisor", *Actualidad en Farmacología y Terapéutica*, vol. 6, no. 1, pp. 50-56.

Hierro-Pescador, J. S. (1997) "Problemas del empirismo en la filosofía de la mente", *Teorema*, vol. 16, no. 2, pp. 35-49.

Hierro-Pescador, J. S. (2005) *Filosofía de la mente y de la Ciencia cognitiva*. Madrid: Akal.

Hierro-Pescador, J. (2006) "¿Por qué hablar de la mente?", *Revista de Filosofía*, vol. 31 no. 2, pp. 67-81.

Hilgard, E. R. (1980) "Consciousness in contemporary psychology", *Annual Review of Psychology*, vol. 31, pp. 1-26.

Hill, C. S. & McLaughlin, B. P. (1998) "There are fewer things in reality than are dreamt of in Chalmer's philosophy", *Philosophy and Phenomenological Research*, vol. 59, no. 2, pp. 445-454.

Hintikka, K. J. J. (1969) “On the logic of perception”, en N. S. Care & R. H. Grimm (eds.), *Perception and Personal Identity*, Cleveland, OH: Case Western Reserve University Press.

Hirstein, W. & Ramachandran, V. S. (1997) “Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of person”, *Proceedings of the Royal Society, Series B: Biological Sciences*, vol. 264, no. 1380, pp. 437-444.

Hobson, J. A. (2001) *The Dream Drugstore: Chemically Altered States of Consciousness*. Cambridge, MA: MIT Press. [Trad. de J. L. Cantero & M. Atienza, Barcelona: Ariel, 2003].

Hobson, J. A. & Fristonb, K. J. (2012) “Waking and dreaming consciousness: Neurobiological and functional considerations”, *Progress in Neurobiology*, vol. 98, no. 1, pp. 82-98.

Holland, O. & Goodman, R. (2003) “Robots with internalised models”, *Journal of Consciosness Studies*, vol. 10, nos. 4-5, pp. 77-109.

Horgan, J. (1996) *The End of Science: Facing the Limits of Science in the Twilight of the Scientific Age*. New York: Broadway Books. [Trad. de B. Moreno Carrillo, Barcelona: Paidós, 1998].

Horgan, J. (2000) *The Undiscovered Mind: How the Human Brain Defies Replication, Medication, and Explanation*. Toucheston-Simon & Schuster. [Trad. de B. Moreno Carrillo, Barcelona: Paidós, 2001].

Horgan, T. (1992) “From cognitive science to folk psychology: Computation, mental representation, and belief”, *Philosophy and Phenomenological Research*, vol. 52, no. 2, pp. 449-484.

Horgan, T. (1994) “Computation and mental representation”, en S. P. Stich & T. A. Warfield (eds), *Mental Representation: A Reader*, Oxford: Blackwell, pp. 302-311.

Horgan, T. & Tienson, J. (2002) “The phenomenology of intentionality and the intentionality of phenomenology”, en D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Reading*, Oxford, UK: Oxford University Press, pp. 520-533.

Houdé, O. (ed.) (2004) *Dictionary of Cognitive Science: Neuroscience, Psychology, Artificial Intelligence, Linguistics, and Philosophy*. New York: Psychology Press.

Hume, D. (1738) *A Treatise of Human Nature. An Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. London: John Noon. [Ed. actual de D. F. Norton & M. J. Norton, Oxford, UK: Oxford University Press, 2000. Trad., introducción y notas de F. Duque Pajuelo, Madrid: Tecnos, 2005].

Humphrey, N. K. (1983) *Consciousness Regained: Chapters in the Development of Mind*. Oxford, UK: Oxford University Press.

Humphrey, N. K. (1987) *The Uses of Consciousness: LVII James Arthur Lecture on the Evolution of the Human Brain*. New York: American Museum of Natural History. [Reimpreso en N. K. Humphrey (ed.), *The Mind Made Flesh: Consciousness and the Physical World*, Oxford, UK: Oxford University Press, 2002, pp. 65-85].

Hurlbert, R. T. & Heavey, C.L. (2001) "Telling what we know: describing inner experience", *Trends in Cognitive Sciences*, vol. 5, no. 9, pp. 400-403.

Hutto, D. D. (2000) *Beyond Physicalism*. Amsterdam: John Benjamins Publishing.

Hutto, D. D. & Myin, E. (2013) *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: MIT Press.

Huxley, T. H. (1866) *Lessons on Elementary Physiology*. London: Macmillan. [Citamos por la ed. revisada por F. Schiller Lee publicada en New York en 1900: Macmillan; Norwood, MA: Norwood Press, 1902].

Huxley, T. H. (1874) "On the hypothesis that animals are automata, and its history", comunicación para la *British Association for the Advancement of Science* aparecida en *The Fortnightly Review. New Series*, vol. 16, pp. 555–580. [Reimpresa en T. H. Huxley, *Collected Essays, Vol 1*, London: Macmillan, 1898, pp. 199-250 (por donde citamos) y, más recientemente, en G. N. A. Vesey (ed.), *Body and Mind. Readings in Philosophy*, London: George Allen & Unwin, 1964, pp. 134-143].

Jack, A. I. & Roepstorff, A. (2002) "Introspection and cognitive brain mapping: From stimulus response to script-report", *Trends in Cognitive Sciences*, vol. 6, no. 8, pp. 333-339.

Jack, A. I. & Shallice, T. (2001) "Introspective physicalism as an approach to the science of consciousness", *Cognition*, vol. 79, nos. 1-2, pp. 161-196.

Jackendoff, R. (1987) *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Jackson, F. C. (1982) “Epiphenomenal qualia”, *Philosophical Quarterly,* vol. 32, no. 127, pp. 127-136.

Jackson, F. C. (1986) “What Mary didn't know”, *Journal of Philosophy*, vol. 83, no. 5, pp. 291-295.

James, W. (1890) *The Principles of Psychology*. New York: Henry Holt. [Trad. de A. Bárcena, México, DF: Fondo de Cultura Económica, 1989].

James, W. (1892) *Psychology. Briefer Course*. Cambridge, MA: Harvard University Press. [Citamos por la ed. de 1984, publicada por la misma editorial con introducción de Michael M. Sokal].

James, W. (1909) *A Pluralistic Universe: Hibbert Lectures at Manchester College on the Present Situation in Philosophy*. New York: Longmans, Green & Co. [Ed. actual: Rockville: Arc Manor, 2008].

James, W. (1911) “Novelty and causation: The perceptual view”, en W. James, *Some Problems of Philosophy*, London: Longmans, Green & Co, pp. 208-219. [Reimpresión actual del original: Lincoln: University of Nebraska Press, Bison, 1996].

Jaume Rodríguez, A. L. (2012) “¿Qué es la teleosemántica? Una perspectiva cartesiana”, *Factótum*, vol. 9, pp. 129-137.

Jaynes, J. (1979) *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. London: Allen Lane. [Trad. de A. Barcena, México, DF: Fondo de Cultura Económica, 1987].

Jing, J., Gillette, R. & Weiss, K. R. (2009) “Evolving concepts of arousal: Insights from simple model systems”, *Reviews in the Neurosciences*, vol. 20, nos. 5-6, pp. 405-427.

Johnson-Laird, P. N. (1983a) *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. N. (1983b), “A computational analysis of consciousness”, *Cognition and Brain Theory*, vol. 6, no. 4, pp. 499-508.

Johnson-Laird, P. N. (1988) *The Computer and the Mind: An Introduction to Cognitive Science*. Cambridge, MA: Harvard University Press. [Trad. de A. Medina, Barcelona: Paidós, 1990].

Kahn, D., Pace-Schott, E. F. & Hobson, J. A. (1997) "Consciousness in waking and dreaming: the roles of neuronal oscillation and neuromodulation in determining similarities and differences", *Neuroscience*, vol. 78, no. 1, pp. 13-38.

Kandel, E. R. (1998) "A new intellectual framework for psychiatry", *The American Journal of Psychiatry*, vol. 155, no. 4, pp. 457-469.

Kandel, E. R. (2006) *In Search of Memory: The Emergence of a New Science of Mind*. New York: W. W. Norton & Co. [Trad. de E. Marengo, Buenos Aires: Katz, 2007].

Kandel, E. R., Schwartz, J. H. & Jessell, T. M. (eds.) (2000) *Principles of Neural Science* (4th ed.). New York: McGraw-Hill. [Trad. de J. L. Agud Aparicio et al., Madrid: McGraw-Hill/Interamericana, 2001].

Kelso, J. A. S. (1995) *Dynamic Patterns*. Cambridge, MA: MIT Press.

Kenny, A. (1963) *Action, Emotion and Will*. London: Routledge & Kegan Paul.

Kihlstrom, J. F. (1987) "The cognitive unsconsciouss", *Science*, vol. 237, no. 4821, pp. 1445-1452.

Kim, J. (1988a) "The mind-body problem after fifty years", en: A. O'Hear (ed.), *Current Issues in Philosophy of Mind*, Cambridge UK: Cambridge University Press, pp. 3-21. [Trad. de J. Vega Encabo & J. L. Vega Encabo, *Azafea*, vol. 4, 2002, pp. 45-63].

Kim, J. (1988b) "What is 'Naturalized epistemology'?", en J. E. Tomberlin (ed.), *Philosophical Perspectives, 2: Epistemology*, Atascadero, CA: Ridgeview Publishing, pp. 381-406. [Reimpreso en J. Kim, *Supervenience and Mind: Selected Philosophical Essays*, Cambridge, UK: Cambridge University Press, 1993, pp. 216-236].

Kim, J. (1995) "Mental causation in Searle's «biological naturalism»", *Philosophy and Phenomenological Research*, vol. 55, no. 1, pp. 189-194.

Kim, J. (1998) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim, J. (2003) "The American origins of philosophical naturalism", en R. Audi (ed.), *Philosophy in America at the Turn of the Century*, *APA Centennial Supplement*, *Journal of Philosophical Research*, Charlottesville, VA: Philosophy Documentation Center, pp. 83-98.

King, C. (2003) "Chaos, quantum-transactions and consciousness: A biophysical model intentional mind", *NeuroQuantology*, vol. 1, no. 1, pp. 129-162.

King, J. & Pribram, K. H. (eds.) (1995) *Scale in Conscious Experience: Is the Brain too Important to be Left to Specialists to Study?* Mahwah, NJ: Lawrence Erlbaum Associates.

Kirk, R. (1974a) "Sentience and behaviour", *Mind*, vol. 83, no. 1, pp. 43-60.

Kirk, R. (1974b) "Zombies vs. materialists", *Proceedings of the Aristotelian Society*, vol. 48 (suppl.), pp. 135-152.

Kirk, R. (1994) *Raw Feeling: A Philosophical Account of the Essence of Consciousness*. Oxford, UK: Oxford University Press.

Kirk, R. (1999) "Why there couldn't be zombies", *Proceedings of the Aristotelian Society*, vol. 73, no. 1 (suppl.), pp. 1-16.

Kirk, R. (2005) *Zombies and Consciousness*. Oxford, UK: Oxford University.

Kirsch, I., Lynn, S. J., Vigorito, M. & Miller, R. R. (2004) "The role of cognition in classical and operant conditioning", *Journal Of Clinical Psychology*, vol. 60, no. 4, pp. 369-392.

Koch, C. (1993) "When looking is not seeing: Towards a neurobiological view of awareness". *Engineering and Science*, vol. 56, no. 3, pp. 2-13.

Koch, C. (2012) *Consciousness. Confessions of a Romantic Reductionist*. Cambridge, MA: MIT Press.

Koch, C. (2014) "A brain structure looking for a function", *Scientific American Mind*, vol. 25, no. 6, pp. 24-27.

Koch, C. & Tsuchiya, N. (2006) "Attention and consciousness: two distinct brain processes", *Trends in Cognitive Sciences*, vol.11, no.1, pp. 16-22.

Koubeissi, M. Z., Bartolomei, F., Beltagyd, A. & Picard, F. (2014) "Electrical stimulation of a small brain area reversibly disrupts consciousness", Epilepsy & Behavior, vol. 37, pp. 32-35.

Koudier, S. (2009) "Neurobiological theories of consciousness", en W. Banks (ed.), *Encyclopedia of Consciousness* (vol. 2), Oxford: Elsevier, pp. 87-100.

Kraut, R. (1982) "Sensory states and sensory objects", *Noûs*, vol. 16, pp. 277-295.

Kriegel, U. (2007) "Philosophical theories of consciousness: Contemporary western perspectives", en P. D. Zelazo, M. Moscovitch & E. Thompson (eds.), *The Cambridge Handbook of Consciousness*, New York: Cambridge University Press, pp. 35-66.

Kriegel, U. (2009) "Mysterianism", en T. Bayne, A. Cleeremans & P. Wilken (eds.), *The Oxford Companion to Consciousness*, Oxford, UK: Oxford University Press, pp. 461-462.

Kriegel, U. & Williford, K. (eds.) (2006) *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press.

Kripke, S. A. (1980) *Naming and Necessity*. Cambridge, MA: Harvard University Press [Trad. de M. M. Valdés, 2ª ed. revisada, México, DF: UNAM, 1995].

Lakatos, Imre. (1970) "History of science and its rational reconstructions", en R. C. Buck y R. S. Cohen (comps.), *Boston Studies in the Philosophy of Science, vol. 8*, Dordrecht: Reidel, pp. 91-108. [Reimpreso en en J. Worrall & G. Currie (eds.), *Irme Lakatos: Philosophical Papers: Vol. 1: The Methodology of Scientific Research Programmes*, New York: Cambridge University Press, 1978, pp. 102-138].

Lamme, V. A. F. (2000) "Neural mechanisms of visual awareness: A linking proposition", *Brain and Mind*, vol. 1, no. 3, pp. 385-406.

Lamme, V. A. F. (2003) "Why visual attention and awareness are different", *Trends in Cognitive Sciences*, vol. 7, no. 1, pp. 12-18

Lamme, V. A. F. (2006) "Towards a true neural stance on consciousness", *Trends in Cognitive Sciences*, vol. 10, no. 11, pp. 494-501.

Lamme, V. A. F. & Roelfsema, P. R. (2000) "The distinct modes of vision offered by feedforward and recurrent processing", *Trends in Neurosciences*, vol. 23, no. 11, pp. 571-579.

Lamote de Grignon, C. (2005) *Antropología neurofilosófica: Un estudio radical de la conducta humana desde los automatismos neonatales al pensar reflexivo del adulto*. Barcelona: Reverté

Lane, R. D., & Nadel, L. (eds.) (2002) *Cognitive Neuroscience of Emotion*. New York: Oxford University Press.

Lange, M. (ed.) (2007) *Philosophy of Science: An Anthology*. Oxford: Blackwell Publishing.

Långsjö, J. W, Alkire, M. T., Kaskinoro, K., Hayama, H., Maksimow, A., Kaisti, K. K, et al. (2012) "Returning from oblivion: Imaging the neural core of consciousness", *The Journal of Neuroscience*, vol. 32, no. 14, pp. 4935-4943.

Laughlin, C. D., McManus, J. & d'Aquili, E. G. (1990) *Brain, Symbol and Experience: Toward a Neurophenomenology of Consciousness*. New York: Columbia University Press.

Laureys, S. Owen, A. M. & Schiff, N. D. (2004) "Brain function in coma, vegetative state, and related disorders", *The Lancet Neurology*, vol. 3, no. 9, pp. 537-546.

Le Van Quyen, M. & Petitmengin, C. (2002) "Neuronal dynamics and conscious experience: an example of reciprocal causation before epileptic seizures", *Phenomenology and the Cognitive Sciences*, vol. 1, no. 2, pp. 169-180.

Le Van Quyen, M. (2003) "Disentangling the dynamic core: A research program for a neurodynamics at the large-scale", *Biological Research*, vol. 36, no. 1, pp. 67-88.

Le Van Quyen, M. (2010) "Neurodynamics and phenomenology in mutual enlightenment: The example of the epileptic aura", en J. Stewart, O. Gapenne & E. A. Di Paolo (eds.), *Enaction: Toward a new paradigm for cognitive science*, Cambridge, MA: The MIT Press, pp. 245-266.

Leibniz, G. W. (1714) *La Monadologie*. [Texto fechado en 1714 y redactado en francés aunque publicado por primera vez en 1720 y en alemán. Citamos por la ed. trilingüe con introducción de Gustavo Bueno. Trad. de J. Velarde, Oviedo: Pentalfa, 1981].

Levine, J. (1983) "Materialism and qualia: The explanatory gap", *Pacific Philosophical Quarterly*, vol. 64, no. 4, pp. 354-361.

Levine, J. (1993) "On leaving out what it's like", en M. Davies & G. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell, pp. 121-136.

Levine, J. (1994) "Out of the closet: A qualophile confronts qualophobia", *Philosophical Topics*, vol. 22, nos. 1-2, pp. 107-126.

Levine, J. (1995) "Qualia: Intrinsic, relational or what?", en T. Metzinger (ed.), *Conscious experience*, Paderborn: Ferdinand Schöningh/Imprint Academic, pp. 277-292.

Levine, J. (1997) "Consciousness located: You'll wonder where the yellow went", *Psycholoquy*, vol. 8, no. 4. [Reseña de Hardcastle, V. G. (1995) *Locating Consciousness*. Amsterdam: John Benjamins].
URL: <http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?locating-consciousness.3>.

Levine, J. (1998) "Conceivability and the metaphysics of mind", *Noûs*, vol. 32. no. 4, pp. 449-480.

Levine, J. (1999) “Conceivability, identity, and the explanatory gap”, en S. R. Hameroff, A. W. Kaszniak, & D. J. Chalmers, (eds.), *Toward a Science of Consciousness III*, Cambridge, MA: MIT Press, pp. 3-12.

Levine, J. (2001) *Purple Haze. The Puzzle of Consciousness*. Oxford, UK: Oxford University Press.

Levine, J. (2007) “Anti-materialist arguments and influential replies”, en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 371-380.

Levitin, D. J. (2002) “Preface”, en D. J. Levitin (ed.) *Foundations of Cognitive Psychology. Core Readings*, Cambridge, MA: MIT Press, pp. xiii-xvi.

Lewis, C. I. (1929) *Mind and the World Order. Outline of a Theory of Knowledge*. New York: Dover Publications. [Reimpreso en 1990].

Lewis, D. (1966) “An Argument for the Identity Theory”, Journal of Philosophy, vol. 63, pp 17–25.

Lewis, D. (1983a) “Individuation by acquaintance and by stipulation”, *Philosophical Review*, vol. 92, pp. 3-32.

Lewis, D. (1983b) “Extrinsic properties”, *Philosophical Studies*, vol. 44, no. 2, pp. 197-200.

Lewis, D. (1983c) “Postscript to ‘Mad pain and martian pain’”, en D. Lewis (ed.) *Philosophical Papers*, vol. 1, Cambridge, UK: Cambridge University Press, pp. 130-132.

Lewis, D. (1986) *On the Plurality of Worlds*. New York, Blackwell.

Lewontin, R. C. (1997) “Billions and billions of demons”, *New York Review of Books*, vol. 44, no. 1, pp. 28-32.

Linton, R. (1936) *The Study of Man: An Introduction*. New York: Appleton Century. [Trad. de D. F. Rubín de la Borbolla, México, DF: Fondo de Cultura Económica, 2006].

Livingston, P. M. (2004) *Philosophical History and the Problem of Consciousness*. Cambridge, UK: Cambridge University Press.

Llinás, R. R. (1987) “’Mindness’ as a functional state of the brain”, en C. Blakemore & S. A. Greenfield (eds.), *Mindwaves: Thoughts on Intelligence, Identity, and Consciousness*, Oxford: Blackwell, pp. 339-358.

Llinás, R. R. (1990) “Intrinsic electrical properties of mammalian neurons and CNS function”, en J-P. Changeux, R. R. Llinás, D. Purves & F. F. Bloom (eds.) *Fidia*

*Research Foundation Neuroscience Award Lectures*, vol. 4. New York: Raven Press, pp. 175-194.

Llinás, R. R. (2001) *I of the Vortex: From Neurons to Self*. Cambridge, MA: MIT Press. [Trad. de E. Guzmán, Bogotá: Norma, 2003].

Llinás, R. R. & Paré, D. (1991) "Of dreaming and wakefulness", *Neuroscience*, vol. 44, no. 3, pp. 521-535.

Llinás, R. R. & Ribary, U. (1993) "Coherent 40-Hz oscillation characterizes dream state in humans", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 5, pp. 2078-2081.

Loar, B. (1981) *Mind and Meaning*. Cambridge, UK: Cambridge University Press.

Loar, B. (1990) ''Phenomenal states'', en J. E. Tomberlin (ed.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, Northridge: Ridgeview Publishing Company, pp. 81-108. [Reimpreso en N. Block, O. Flanagan, & G. Güzeldere (eds.), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, 1997, pp. 597-616].

Locke, J. (1690) *An Essay Concerning Human Understanding* [Trad. de E. O'Gorman, México, DF: Fondo de Cultura Económica, 2005].

Lormand, E. (1995) "Qualia! (Now showing at a theater near you)", *Philosophical Topics*, vol. 22, nos. 1-2, pp. 127-156.

Lormand, E. (1998) "Consciousness", en E. Craig & L. Floridi (eds.), *Routledge Encyclopedia of Philosophy, Version 1.0. Philosophy of Mind*, London: Routledge, pp. 109-124.

Ludwig, K. (1993) "A dilemma for searle's argument for the connection principle", *The Behavioral and Brain Sciences*, vol. 16, no. 1, pp. 194-195.

Lukomski, A (2007) "El problema mente-cuerpo", *Logos. Revista de Filosofía de la Facultad de Filosofía y Letras*, no. 12, pp. 57-68.

Luria, A. R. (1962) *Высшие корковые функции человека и их нарушения при локальных поражениях мозга*. Москва: Издательство Московского университета. [Primera ed. inglesa: *Higher Cortical Functions in Man*, Nueva York: Basic Books, 1966. Trad. española de A. Parés, Barcelona: Fontanella, 1983].

Lutz, A. (2002) "Toward a neurophenomenology of generative passages: A first empirical case study", *Phenomenology and the Cognitive Sciences*, vol. 1, no. 2, pp. 133-167.

Lutz, A., Lachaux, J-P, Martinerie, J., & Varela, F. (2002) "Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 3. pp. 1586-1591.

Lutz, A. & Thompson, E. (2003) "Neurophenomenology. Integrating subjective experience and brain dynamics in the neuroscience of consciousness", *Journal of Consciousness Studies*, vol. 10, nos. 9-10, pp. 31-52.

Lycan, W. G. (1987) *Consciousness*. Cambridge, MA: Bradford Books/MIT Press.

Lycan, W. G. (1995) "Consciousness as internal monitoring", en en J. E. Tomberlin (ed.), *Philosophical Perspectives, 9: AI, Connectionism, and Philosophical Psychology*, Atascadero, CA: Ridgeview Publishing, pp. 1-14. [Reimpreso en N. Block, O. Flanagan, & G. Güzeldere (eds.), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, 1997, pp. 755-771].

Lycan, W. G. (1996) *Consciousness and Experience*. Cambridge, MA: Bradford Books/MIT Press.

Lycan, W. G. (2001) "The case for phenomenal externalism", en J. E. Tomberlin (ed.), *Philosophical Perspectives, 15: Metaphysics*, Atascadero, CA: Ridgeview Publishing, pp. 17-35.

Lycan, W. (2006a) "Representational theories of consciousness", en E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Autumn 2006 Edition)*. URL: <http://plato.stanford.edu/entries/consciousness-representational/>.

Lycan, W. (2006b) "Resisting ?-ism", *Journal of Consciousness Studies*, vol. 13, nos. 10-11, pp. 65-71. [Reimpreso en A. Freeman (ed.), *Consciousness and its Place in Nature: Does Physicalism entail Panpsychism?*, Exeter: Imprint Academic, 2006, pp. 65-71].

MacCormac, E. & Stamenov, M. I. (eds.) (1996) *Fractals of Brain, Fractals of Mind: In Search of a Symmetry Bond*. Philadelphia: John Benjamins.

Macphail, E. M. (2000) "The search for a mental rubicon", en C. M. Heyes & L. Huber (eds.), *The Evolution of Cognition*, Cambridge, MA: MIT Press., pp. 253-271.

Maddy, P. (2007) *Second Philosophy. A Naturalistic Method*. New York: Oxford University Press.

Manns, J. R & Buffalo, E. A. (2013) "Learning and memory: Brain systems", en L. R. Squire, D. Berg, F. E. Bloom, S. du Lac, A. Ghosh & N. C. Spitzer (eds.), *Fundamental Neuroscience* (4th ed.), Elsevier: New York, pp. 1029-1051.

Marcel, A. J. & Bisiach, E. (eds.) (1988) *Consciousness in Contemporary Science*. Oxford, UK: Clarendon Press.

Marcos Malmierca, J. L. (2014) "Evidencia de aprendizaje inconsciente generado mediante priming asociativo enmascarado", *Psicológica*, vol. 35, pp. 291-308.

Margulis, L. & Sagan, D (2002) *Acquiring Genomes: A Theory of the Origins of Species*. New York: Basic Books. [Trad. de D. Sempau, Barcelona: Kairós, 2003].

Martínez, M. (2008) "La 'P' de PANIC. Representacionalismo y fenomenología del dolor", *Teorema*, vol. 27, no. 3, pp. 181-195.

Martín-Loeches, M. (2008) *La mente del 'Homo sapiens'. El cerebro y la evolución humana*. Madrid: Aguilar-Santillana.

Martín-Rodríguez, J. F., Cardoso-Pereira, N., Bonifacio, V. & Barroso y Martín, J. M. (2004) "La década del cerebro (1990-2000): algunas aportaciones", *Revista Española de Neuropsicología*, vol. 6, nos. 3-4, pp. 131-170.

Martínez-Freire, P. F. (1992) "Delimitación de las ciencias cognitivas", *Logos. Anales del Seminario de Metafísica*, no. extra 1, pp. 443-452.

Martínez-Freire, P. F. (1995) *La nueva filosofía de la mente*. Barcelona: Gedisa.

Martínez-Freire, P. F. (1996) "La filosofía de la mente, hoy", *Revista de Historia de la Psicología*, vol. 17, nos. 3-4, pp. 299-304.

Martínez-Freire, P. F. (1999) "El debate mente-cerebro a la luz de las nuevas técnicas de exploración del cerebro", *Contrastes, Revista Interdisciplinar de Filosofía*, vol. 4, pp. 65-75.

Martínez-Freire, P. F. (2007a) *La importancia del conocimiento. Filosofía y ciencias cognitivas* (2ª ed.). A Coruña: Netbiblo.

Martínez-Freire, P. F. (2007b) "Del problema mente-cuerpo al problema mente-cerebro", en A. Segura Naya (ed.), *Historia universal del pensamiento filosófico*, vol. 5, Ortuella, Vizcaya: Liber, pp. 799-811.

Marr, D. (1982) *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman & Co. [Trad. de T. del Amo Martín, Madrid: Alianza, 1985].

Mashour, G. A. & Alkire, M. T. (2013) "Evolution of consciousness: Phylogeny, ontogeny, and emergence from general anesthesia", *Proceedings of the National*

*Academy of Sciences of the United States of America*, vol. 110, suppl. 2, pp. 10357-10364.

Maxwell, N. (1966) "Physics and common sense", *British Journal for the Philosophy of Science*, vol. 16, no. 64, pp. 295-311.

Maxwell, N. (1968) "Understanding sensations", *Australasian Journal of Philosophy*, vol. 46, no. 2, pp. 127-146.

Maxwell, N. (2011) "Three philosophical problems about consciousness and their possible resolution", *Open Journal of Philosophy*, vol. 1, no. 1, pp. 1-10.

Maynard Smith, J. (1998) *Shaping Life. Genes, Embryos and Evolution*. London: Weidenfeld & Nicolson. [Trad. de A. J. Desmonts, Barcelona: Crítica, 2000].

Mayr, E. W. (1997) *This is Biology. The Science of the Living World*. Cambridge, MA: Harvard University Press.

McCarthy, J. & Hayes, P. J. (1969) "Some philosophical problems from the standpoint of artificial intelligence", en B. Meltzer & D. M. Michie (eds.), *Machine Intelligence. Vol. 4*, Edinburgh: Edinburgh University Press, pp. 463-502.

McClelland, J. L., Rumelhart, D. E. & the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2*. Cambridge, MA: MIT Press.

McDermott, D. V. (2001) *Mind and Mechanism*. Cambridge, MA: MIT Press.

McGinn, C. (1982/1996) *The Character of Mind: An Introduction to the Philosophy of Mind* (2nd ed.). New York: Oxford University Press.

McGinn, C. (1989) "Can we solve the mind-body problem?", *Mind, New Series*, vol. 98, no. 391, pp. 349-366.

McGinn, C. (1991) *The Problem of Consciousness: Essays Toward a Resolution*. Oxford: Blackwell.

McGinn, C. (1993) *Problems in Philosophy: The Limits of Inquiry*. Oxford: Blackwell.

McGinn, C. (2002) *The Making of a Philosopher: My Journey Through Twentieth-Century Philosophy*. New York: Harper & Collins.

McGinn, C. (2004) *Consciousness and its Objects*. New York: Oxford University Press.

McGinn, C. (2012) "All machine and no ghost?", *New Statesman*, 20 Feb. 2012. URL: <http://www.newstatesman.com/ideas/2012/02/consciousness-mind-brain>.

McMullen, C. (1985) "'Knowing 'what it's like' and the essential indexical", *Philosophical Studies*, vol. 48, no. 2, pp. 211-233.

Menzel, R. (2009) “Serial position learning in honeybees”, *PLoS ONE*, vol. 4, no. 3. DOI: <10.1371/journal.pone.0004694>.

Menzel, R., Greggers, U., Smith, A., Berger, S. Brandt, R., Brunke, S., et al. (2005) “Honey bees navigate according to a map-like spatial memory”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 8, pp. 3040-3045.

Merleau-Ponty, M. (1945) *Phénoménologie de la Perception*. Paris: Gallimard. [Trad. de J. Cabanes, Barcelona: Península, 2000].

Miklos, G. L. G. (1998) “The evolution and modification of brains and sensory systems”, *Daedalus*, vol. 127, no. 2, pp. 197-216.

Miller, G. A. (2003) “The cognitive revolution: A historical perspective”, *Trends in Cognitive Sciences*, vol. 7, no. 3, pp. 141-144.

Millikan, R. G. (2001) “What has natural information to do with intentional representation?”, en D. Walsh (ed.), *Evolution, Naturalism and Mind*, Cambridge, UK: Cambridge University Press, pp. 105-126.

Milner, A. D. & Goodale, M. A. (2006) *The Visual Brain in Action* (2nd ed.). Oxford, UK: Oxford University Press.

Minsky, M., & Papert, S. (1969) *Perceptrons*. Cambridge, MA: MIT Press.

Mitchell, M. (2009) *Complexity: A Guided Tour*. New York: Oxford University Press.

Montecucco, L. (2002) “Can supervenience save the mental?”, en E. Agazzi & L. Montecucco (eds.), *Complexity and Emergence*, Singapore: World Scientific Publishing, pp. 161-180.

Montesinos Sierra, J. (2004) “La matematización de la naturaleza como vía única de la ciencia”, *Actas del Seminario Orotava de Historia de la Ciencia. Los Orígenes de la Ciencia Moderna*, años XI-XII, pp. 351-369.

Mora Teruel, F. (2001) *El reloj de la sabiduría. Tiempos y espacios en el cerebro humano*. Madrid: Alianza.

Mora Teruel, F. (2004) *Genios, locos y perversos. Cerebro, enfermedad mental y diversidad humana*. Madrid: Alianza.

Mora Teruel, F. (2007) *Neurocultura. Una cultura basada en el cerebro*. Madrid: Alianza.

Mora Teruel, F. (2013) “¿Qué es una emoción?”, *Arbor*, vol. 189, no. 759, a004. <doi: http://dx.doi.org/10.3989/arbor.2013.759n1003>.

Moreland, J. P. (2008) *Consciousness and the Existence of God. A Theistic Argument*. New York: Routledge.

Morgan, C. L. (1923) *Emergent Evolution. The Gifford Lectures delivered in the University of St. Andrews in the Year 1922*. London: Williams and Norgate.

Morin, E. (1987) *Penser l'Europe*. Paris: Gallimard. [Trad. de B. E. Anastasi de Lonné, Barcelona: Gedisa, 2003].

Moruzzi, G. & Magoun, H. W. (1949) "Brain stem reticular formation and activation of the EEG", *Electroencephalography and Clinical Neurophysiology*, vol. 1, no. 4, pp. 455-473.

Mosterín, J. (1984/2000) *Conceptos y teorías en la ciencia* (3ª ed.). Madrid: Alianza.

Mosterín, J. (2003a) "El espejo roto del conocimiento y el ideal de una visión coherente del mundo", *Revista Iberoamericana de Ciencia, Tecnología y Sociedad*, vol. 1, no. 1, pp. 209-221.

Mosterín, J. (2003b) "La insuficiencia de los paradigmas metafóricos en psicología", *Revista de la Asociación Española de Neuropsiquiatría*, vol. 23, no. 85, pp. 89-104.

Mosterín, J. (2006) *Aristóteles*. Madrid: Alianza.

Mosterín, J. (2006/2008) *La naturaleza humana*. Madrid: Espasa Calpe.

Mosterín, J. (2007) *Los lógicos*. Madrid: Espasa Calpe.

Mosterín, J. (2013a) *Ciencia, filosofía y racionalidad*. Barcelona: Gedisa.

Mosterín, J. (2013b) *El reino de los animales*. Madrid: Alianza.

Moural, J. (2003) "The Chinese room argument", en B. Smith (ed.), *John Searle*, Cambridge, UK: Cambridge University Press, pp. 214-260.

Moutoussis, K. & Zeki, S. (1997) "Functional segregation and temporal hierarchy of the visual perceptive systems", *Proceedings of the Royal Society, Series B: Biological Sciences*, vol. 264, no. 1387, pp. 1407-1414.

Moya, A. (2009) "Biología de la vida y la conciencia. A propósito de Darwin", *Ludus Vitalis*, vol. 17, no. 32, pp. 385-394.

Moya, A. (2010) *Pensar desde la ciencia*. Madrid: Trotta.

Moya, C. J. (2004) *Filosofía de la mente*. Valencia: Universitat de València.

Moya Santoyo, J. (1999) "La recuperación de la conciencia en la ciencia cognitiva. Un estudio a través de Psycinfo & Psyclit (1994-1998)", *Revista de Historia de la Psicología*, vol. 24, nos. 3-4, pp. 197-208.

Murphy, A. E. (1945) "Review of *Naturalism and the Human Spirit*", *Journal of Philosophy*, vol. 42, no. 15, pp. 400-417.

Nagel, T. (1974) "What is it like to be a bat?", *The Philosophical Review*, vol. 83, no. 4, pp. 435-450.

Nagel, T. (1986) *The View from Nowhere*. New York: Oxford University Press.

Nagel, T. (1993a) "The mind wins!", *The New York Review of Books*, vol. 40, no. 5, pp. 37-41.

Nagel, T. (1993b) "What is the mind-body problem?", en G. R. Bok and J. Marsh (eds.), *Experimental and Theoretical Studies of Consciousness: Ciba Foundation Symposium, 174*, Chichester, UK: Wiley, pp. 1-13.

Nagel, T. (1998) "Conceiving the impossible and the mind-body problem", *Philosophy*, vol. 73, no. 285, pp. 337-352.

Nagel, T. (2000) "The psychophysical nexus", en P. A. Boghossian & C. Peacocke (eds.), *New Essays on the A Priori*, New York: Oxford University Press, pp. 433-471.

Nagel, T. (2012) *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is almost Certainly False*. New York: Oxford University Press.

Nelkin, N. (1996). "Subjectivity", en S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Cambridge, MA: Blackwell, p. 568.

Neumann, O. & Klotz, W. (1994) "Motor responses to nonreportable, masked stimuli: Where is the limit of direct parameter specification?", en C. Umilta and M. Moscovitch (eds.), *Attention and Performance XV. Conscious and Nonconscious Information Processing*, Cambridge, MA: MIT Press. pp. 123-150.

Newell, A. & Simon, H. A. (1972) *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Nida-Rümelin, M. (2002) "Causal reduction, ontological reduction, and first-person ontology. Notes on Searle's views about consciousness", en G. Grewendorf & G. Meggle (eds.), *Speechs Acts, Mind, and Social Reality. Discussions with John R. Searle*, Dordrecht: Kluwer Academic Publishers, pp. 205-221.

Nisbet, R. A. (1970) *The Social Bond: An Introduction to the Study of Society*. New York: Alfred A. Knopf. [Trad. de M. Bouyat, Barcelona: Vicens-Vives, 1975].

Noë, A. (2004) *Action in Perception*. Cambridge, MA: MIT Press.

Noë, A. (2009) *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York: Hill & Wang. [Trad. de N. d' Amonville Alegría, Barcelona: Kairós, 2010].

Nordby, K. (1990) "Vision in a complete achromat: a personal account", en R. F. Hess, L. T. Sharpe & K. Nordby (eds.), *Night Vision: Basic, Clinical and Applied Aspects*, Cambridge, UK: Cambridge University Press, pp. 290-315.

Norman, D. A. & Shallice, T. (1980) "Attention to action: Willed and automatic control of behaviour", *Center for Human Information Processing Technical Report*, no. 99. [Reimpreso en R. Davidson, G. Schwartz & D. Shapiro, (eds.), *Consciousness and Self Regulation: Advances in Research and Theory Vol. 4*, New York: Plenum, 1986, pp. 1-18].

Nunziante, A. M. (2013) "The «Morbid fear of the subjective»", *Metodo. International Studies in Phenomenology and Philosophy*, vol. 1, no. 2, pp. 39-57.

O'Connor, T. & Wong, H. Y. (2002) "Emergent properties", en E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2005 Edition)*. URL: <http://plato.stanford.edu/archives/sum2005/entries/propertiesemergent/>.

O'Dea, J. (2002) "The indexical nature of sensory concepts", *Philosophical Papers*, vol. 31, no. 2, pp. 169-181.

O'Regan, J. K. & Noë, A. (2001) "A sensorimotor approach to vision and visual consciousness", *Behavioral and Brain Sciences*, vol. 24, no. 5, pp. 939-973.

O'Shea, M. (2005) *The Brain. A Very Short Introduction*. New York: Oxford University Press.

Ojeda, C. (2001) "Francisco Varela y las ciencias cognitivas", *Revista Chilena de Neuro-Psiquiatría*, vol. 39, no. 4, pp. 286-295.

Olafson, F. (1994) "Brain dualism", *Inquiry*, no. 37, vol. 2, pp. 253-265.

Oppenheim, P. & Putnam, H. (1958) "Unity of science as a working hypothesis", en H. Feigl, M. Scriven & G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science, Volume II*, Minneapolis: University of Minnesota Press, pp 3-36. [Reimpreso en R. Boyd, P. Gasper, J. D. Trout (eds.), *The Philosophy of Science*, Cambridge, MA: MIT Press, 1991, pp. 405-427].

Ornstein, R. E. (1972/1986) *The Psychology of Consciousness* (3rd ed.). New York: Penguin.

Ortega Escobar, J. (2014) "Evolución del cerebro y la cognición", en P. Enríquez de Valenzuela (coord.), *Neurociencia cognitiva*, Madrid: Sanz y Torres, pp. 57-74.

Palmer, D. E. (1998) “Searle on consciousness: or How not to be a physicalist”, *Ratio*, vol. 11, no. 2, pp. 159-169.

Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press.

Panksepp, J. (2011) “Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals”, *PLoS ONE*, vol. 6, no. 9. DOI: <10.1371/journal.pone.0021236>.

Panksepp, J. (2012) “Do Animals have Emotional Lives?”, en C. Jakobsson (ed.), *Sustainable Agriculture*, Uppsala: Baltic University Press, pp. 316-323.

Panksepp, J. & Biven, L. (2012) *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. New York: W. W. Norton & Co.

Papineau, D. (1993) *Philosophical Naturalism*. Oxford: Blackwell.

Papineau, D. (1995) “The antipathetic fallacy and the boundaries of consciousness”, en T. Metzinger, (ed.), *Conscious Experience*. Paderborn: Ferdinand Schöningh, pp. 259-272.

Papineau, D. (2002) *Thinking about Consciousness*. Oxford, UK: Oxford University Press.

Papineau, D. (2006) “Phenomenal and perceptual concepts”, en T. Alter & S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*, New York: Oxford University Press, pp. 111-144.

Papineau, D. (2007) “Naturalism”, en E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy (Spring 2007 Edition)*.
URL: <http://plato.stanford.edu/entries/naturalism/>.

Papineau, D. (2011) “What exactly is the explanatory gap?”, *Philosophia*, vol. 39, no. 1, pp. 5-19.

Papini, M. R. (1999) “Problemas y enfoques de la psicología comparada del aprendizaje”, *Revista Mexicana de Análisis de la Conducta*, vol. 25, no. 2, pp. 217-235.

Paré, D. & Llinás, R. (1995) “Conscious and pre-conscious processes as seen from the standpoint of sleep-waking cycle neurophysiology”, *Neuropsychologia*, vol. 33, no. 9, pp. 1155-1168.

Parker, A. E., Wilding, E. L. & Bussey, T. J. (eds.) (2002) *The Cognitive Neuroscience of Memory: Episodic Encoding and Retrieval*. Hove: Psychology Press.

Parvizi, J. & Damasio, A. R. (2001) "Consciousness and the brainstem", *Cognition*, vol. 79, nos. 1-2, pp. 135-159.

Pawlik, K. (1998) "The neuropsychology of consciousness: The mind-body problem re-addressed", *International Journal of Psychology*, vol. 33, no. 3, pp. 185-189.

Peacocke, C. (1983) *Sense and Content: Experience, Thought and their Relations*. Oxford, UK: Oxford University Press.

Peacocke, C. (2001) "Does perception have a nonconceptual content?", *The Journal of Philosohy*, vol. 98, no. 5, pp. 239-264.

Pearson, M. P. (1999) *The Archaeology of Death and Burial*. College Station, Texas: Texas A&M Press.

Penrose, R. (1989) *The Emperor's New Mind: Computers, Minds and the Laws of Physics*, Oxford, UK: Oxford University Press. [Trad. de J. J. García Sanz, México, DF: Fondo de Cultura Económica, 1996].

Pereboom, D. (1994) "Bats, brain scientists, and the limitation of introspection", *Philosophy and Phenomenological Research* , vol. 54, no. 2, pp. 315-329.

Pereira, G. (1554) *Antoniana Margarita*. Santiago de Compostela: Servicio de publicaciones de la Universidad de Santiago de Compostela. [Texto traducido del latín por J. L. Barreiro Barreiro & C. Souto García y publicado por primera vez en español en el año 2000].

Pérez, D. I. (2002) "Los qualia desde un punto de vista naturalista", *Azafea*, vol. 4, pp. 65-83.

Pérez Chico, D. (1999) "¿Problema? ¿Qué problema? Naturalismo biológico y el problema mente-cuerpo", *Teorema*, vol. 18, no. 1, pp. 125-138.

Perry, J. (2001) *Possibility, Consciousness and Conceivability*. Cambridge, MA: MIT Press.

Perry, J. (2006) "Using indexicals", en M. Devitt & R. Hanley (eds.), *The Blackwell Guide to the Philosophy of Language*, Oxford: Blackwell, pp. 314-344.

Pineda, D. (1999) "Searle y el problema de la exclusión causal. Vindicación del materialismo frente al naturalismo biológico", *Teorema*, vol. 18, no. 1, pp. 155-170.

Pinillos, J. L. (1978) "Lo físico y lo mental", *Boletín Informativo de la Fundación Juan March*, vol. 71, pp. 3-31. [Reimpreso en J. L. Pinillos, *La psicología y el hombre de hoy*. México, DF: Trillas, 1983, pp. 141-166].

Pinker, S. (1997) *How the Mind Works*. New York: W. W. Norton & Co. [Trad. de F. Meler-Orti, Barcelona: Destino, 2001].

Pinker, S. (2002) *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking. [Trad. de R. Filella Escolà, Barcelona: Paidós, 2003].

Pisella, L., Arzi, M. & Rossetti, Y. (1998) "The timing of color and location processing in the motor context", *Experimental Brain Research*, vol. 121, no. 3, pp. 270-276.

Pitkänen, M. (2003) "TGD (topological geometro dynamics) inspired theory of consciousness", *NeuroQuantology*, vol. 1, no. 1, pp. 68-93.

Place, U. T. (1956) "Is consciousness a brain process?", *British Journal of Psychology*, vol. 47, no. 1, pp. 44-50.

Platón (ca. 386 a. C.) *Gorgias*, en *Diálogos II: Gorgias, Menéxeno, Eutidemo, Menón, Crátilo*, Madrid: Gredos, 2004. [Trad. de J. Calonge, E. Acosta, F. Oliveri & J. Calvo].

Platón (ca. 370 a. C.) *Fedro*, en *Diálogos III: Fedón, Banquete, Fedro*, Madrid: Gredos, 2004. [Trad. de C. García Gual, M. Martínez Hernández & E. Lledó Íñigo].

Platón (ca. 360 a. C.) *Crátilo*, en *Diálogos II: Gorgias, Menéxeno, Eutidemo, Menón, Crátilo*, Madrid: Gredos, 2004. [Trad. de J. Calonge, E. Acosta, F. Oliveri & J. Calvo].

Plazzi, G., Vetrugno, R., Provini, F & Montagna, P. (2005) "Sleepwalking and other ambulatory behaviours during sleep", *Neurological Sciences*, vol. 26, no. 3 (suppl.), pp. 193-198.

Plum, F. (1991) "Coma and related global disturbances of the human conscious state", en A. Peters y E. G. Jones (eds.), *Normal and Altered States of Function. Vol. 9*, New York: Plenum Press, pp. 359-425.

Polger, T. (2007) "Rethinking the evolution of consciousness", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 72-86.

Pöppel, E., Held, R. & Frost, D. (1973) "Residual visual function after brain wounds involving the central visual pathways in man", *Nature*, vol. 243, no. 5405, pp. 295-296.

Popper, K. R. (1934/1959) *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Tübingen: Mohr Siebeck. [Reescrito por el autor en inglés y

publicado en New York: Basic Books, 1959. Trad. española de V. Sánchez de Zavala, Madrid: Tecnos, 2008 (2ª ed.)].

Posner, M. I. (ed.) (2004/2012) *Cognitive Neuroscience of Attention*. New York: Guilford Press.

Posner, M. I. & Boies, S. J. (1971) "Components of attention", *Psychological Review*, vol. 78, no. 5, pp. 391-408.

Posner, M. I. & Snyder, C. R. R. (1975) "Attention and cognitive control", en R. L. Solso (ed.), *Information Processing and Cognition: The Loyola Symposium*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 55-84.

Posner, M. I., & Warren, R. E. (1972) "Traces, concepts, and conscious constructions", en A. W. Melton & E. Martin (eds.), *Coding Processes in Human Memory*, Chichester: Winston & Willey, pp. 25-44.

Pozo, J. I. (2001) *Humana mente. El mundo, la conciencia y la carne*. Madrid: Morata.

Priest, S. (1991) *Theories of the Mind*. New York: Houghton Mifflin [Trad. de. C. García Trevijano & S. Nuccetelli, Madrid: Cátedra, 1994].

Prigogine, I. (1988) *La nascita del tempo*. Roma: Theoria [Trad. de J. M. Pons, Barcelona: Tusquets, 2012].

Prinz, J. (2000) "A neurofunctional theory of visual consciousness", *Consciousness and Cognition*, vol. 9, no. 2, pp. 243-259.

Prinz, J. (2001) "Functionalism, dualism, and the neural correlates of consciousness", en W. Bechtel, P. Mandik, J. Mundale & R. Stufflebeam (eds.), *Philosophy and the Neurosciences: A Reader*, Oxford: Blackwell, pp. 278-294.

Prinz, J. (2007) "The intermediate level theory of consciousness", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 247-260.

Prinz, J. (2009) "Is consciousness embodied?", en P. Robbins & M. Aydede (eds.), *Cambridge Handbook of Situated Cognition*, Cambridge, UK: Cambridge University Press, pp. 419-436.

Proust, J. (1999) "Mind, space, and objectivity in non-human animals", *Erkenntniss*, vol. 51, no. pp. 41-58.

Putnam, H. W. (1967) "Psychological predicates", en W. H. Capitan & D. D. Merrill (eds.), *Art, Mind, and Religion*, Pittsburgh: University of Pittsburgh Press, pp. 37-48. [Reimpreso bajo el título "The nature of mental states" en H. W. Putnam,

*Mind Language and Reality: Philosophical Papers, vol. 2*, Cambridge, UK: Cambridge University Press, 1975, pp. 229-440].

Putnam, H. W. (1975) "The meaning of meaning", en K. Gunderson (ed.), *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science 7*, Minneapolis: University of Minnesota Press, pp. 131-193. [Reimpreso en H. W. Putnam, *Mind Language and Reality: Philosophical Papers, vol. 2*, Cambridge, UK: Cambridge University Press, 1975, pp. 215-271].

Putnam, H. W. (1982) "Why reason can't be naturalized", *Synthese*, vol. 52, no. 1, pp. 3-23. [Reimpreso en H. W. Putnam, *Realism and Reason: Philosophical Papers, vol. 3*, Cambridge, UK: Cambridge University Press, 1983, pp. 229-247].

Quine, W. v. O. (1951) "Two dogmas of empiricism', *Philosophical Review*, vol. 60, pp. 20-43. [Reimpreso en W. v. O. Quine, *From a Logical Point of View* (2nd ed.), Cambridge, MA: Harvard University Press, 1961, pp. 20-46. Trad. de M. Sacristán, Barcelona: Ariel, 1962].

Quine, W. v. O. (1960) *Word and Object*. Cambridge, MA: MIT Press. [Trad. de M. Sacristán, Barcelona: Labor, 1968].

Quine. W. v. O. (1970) *Pursuit of Truth*. Cambridge, MA: Harvard University Press. [Trad. de J. Rodríguez Alcázar, Barcelona: Crítica, 1992].

Quine, W. v. O. (1979) "Has philosophy lost contact with people?", *Long Island Newsday*, Nov. 18, part. 1, sec. 2: 5, p. 13. [Reimpreso en W.v. O. Quine, *Theories and Things*. Cambridge, MA: Harvard University Press, 1981, pp. 190-193, por donde citamos. Trad. de A. Zirión Quijano, México, DF: Instituto de Investigaciones Filosóficas. Universidad Nacional Autónoma de México, 1986].

Rakova, M. (2006) *Philosophy of Mind A-Z*. Edinburgh: Edinburgh University Press.

Ramachandran, V. S. (2004a) *A Brief Tour of Human Consciousness: From Impostor Poodles to Purple Numbers*. New York: Pi Press.

Ramachandran V. S. (2004b) "The astonishing Francis Crick", *Perception*, vol. 33, no. 10, pp. 1151-1154.

Ramachandran, V. S. & Blakeslee, S. (1998) *Phantoms in the brain*. New York: William Morrow and Company.

Ramachandran, V. S. & Hirstein, W. (1997) "Three laws of qualia: What neurology tells us about the biological functions of consciousness, qualia and the self", *Journal of Consciousness Studies*, vol. 4, nos. 5-6, pp. 429-458.

Ramón y Cajal, S. (1888) "Estructura de los centros nerviosos de las aves", *Revista Trimestral de Histología Normal y Patológica*, vol. 1, pp. 1-10. [Reimpreso en S. Ramón y Cajal, *Trabajos escogidos 1880-1890*, Barcelona: Antoni Bosch, 2006, pp. 305-316].

Ramón y Cajal, S. (1891) "Comunicación acerca de la significación fisiológica de las prolongaciones protoplásmicas y nerviosas de las células de la sustancia gris", en F. Barbera (dir.), *Actas y detalles del Iº Congreso Médico-Farmacéutico Regional*, Valencia: Imprenta de F. Domenech, 1894, pp. 70-85. [Previamente publicado en *Revista de Ciencias Médicas de Barcelona*, vol. 17, 1891, pp. 671-679, 715-723].

Ramsey, W. M. (2007) *Representation Reconsidered*. Cambridge, UK: Cambridge University Press.

Reichenbach, H. (1949) "The philosophical significance of the theory of relativity", en P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, La Salle, IL: Open Court, pp. 287-311.

Rensch, B. (1972) *Homo Sapiens. From Man to Demigod*. New York: Columbia University Press. [Trad. de A. Guéra Miralles, Madrid: Alianza, 1980].

Revonsuo, A. (2000) "The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming", *Behavioral and Brain Sciences*, vol. 23, no. 6, 877-901.

Revonsuo, A. (2005) *Inner Presence: Consciousness as a Biological Phenomenon*. Cambridge, MA: MIT Press.

Revonsuo, A. & Valli, K. (2000) "Dreaming and consciousness: Testing the threat simulation theory of the function of dreaming", *Psyche*, vol. 6, no. 8. URL: <http://www.theassc.org/files/assc/2467.pdf>.

Rey, G. (1983) "A reason for doubting the existence of consciousness", en R. Davidson, G. Schwartz & D. Shapiro (eds.), *Consciousness and Self-Regulation, vol 3*, New York: Plenum, pp. 1-39.

Rey, G. (1988) "A question about consciousness", en H. Otto & J. Tuedio (eds.), *Perspectives on Mind*, Dorderecht: Reidel, pp. 5-24.

Rey, G. (1990) "Constituent causation and the reality of mind", *Behavioral and Brain Sciences*, vol. 13, no. 4, pp. 620-621.

Rey, G. (1998) "A narrow representationalist account of cualitative experience", en J. E. Tomberlin (ed.), *Philosophical Perspectives 12, Language, Mind, and Ontology*, Atascadero, CA: Ridgeview Publishing, pp. 435-458.

Rivière, A. (1987) *El sujeto de la psicología cognitiva*. Madrid: Alianza.

Rivière, A. (1991) "Orígenes históricos de la psicología cognitiva: Paradigma simbólico y procesamiento de la información", *Anuario de Psicología*, no. 51, pp. 129-155.

Rodríguez, E., George, N., Lachaux, J-P., Martinerie, J., Renault, B. & Varela, F. J. (1999) "Perception's shadow: long-distance synchronization of human brain activity", *Nature*, vol. 397, no. 6714, pp. 430-433.

Rodríguez González, M. (2006) "¿Es el naturalismo un humanismo? Sobre la 'imposibilidad' de lo mental", *Logos: Anales del seminario de metafísica*, vol. 39, pp. 329-340.

Rodríguez González, M. (2010) "La conciencia fenoménica y el límite del naturalismo", *Convivium*, vol. 23, pp. 173-188.

Roederer, J. G. (2003) "On the concept of information and its role in nature", *Entropy*, vol. 5, no. 1, pp. 3-33.

Romanes, G. J. (1883) *Mental Evolution in Animals*. London: Kegan Paul, Trench & Co. [Ed. actual: Whitefish, Montana: Kessinger Publishing, 2004].

Romanes, G. J. (1892) *Darwin and After Darwin. An Exposition of the Darwinian Theory and a Discussion of Post-Darwinian Questions*. *Vol. I. The Darwinian Theory*. Chicago: The Open Court. [Citamos por la cuarta ed., publicada por la misma editorial en 1910].

Rorty, R. M. (1979) *Philosophy and the Mirror of Nature*. Oxford: Blackwell. [Trad. de J. Fernández Zulaica, Madrid: Cátedra, 1989].

Rose, F. (1985) *Into the Heart of the Mind: An American Quest for Artificial Intelligence*. New York: Vintage Books.

Rosenberg, A. (2011) "Why I am a naturalist", *The New York Times*, Sep. 17. URL: <http://opinionator.blogs.nytimes.com/2011/09/17/why-i-am-a-naturalist/?ref=opinion>.

Rosenblatt, F. (1962) *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Washington, DC: Spartan Books.

Rosenthal, D. M. (1986) "Two concepts of consciousness", *Philosophical Studies*, vol. 49, no. 3, pp. 329-359.

Rosenthal, D. M. (1990) *A theory of consciousness*. Report no. 40. Center for Interdisciplinary Research, Research Group on Mind and Brain, Bielefeld: University of Bielefeld. [Reimpreso en N. Block, O. Flanagan, & G. Güzeldere (eds.), *The Nature of Consciousness. Philosophical Debates*, Cambridge, MA: MIT Press, 1997, pp. 729-753].

Rosenzweig, M. R., Breedlove, S. M. & Watson, N. V. (1999/2005) *Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience* (4th ed.). Sunderland, MA: Sinauer. [Trad. de la segunda ed. de J. Soler, Barcelona: Ariel, 2001].

Ross, J. A. (2008) "Hitting on consciousness: Honderich versus McGinn", *Journal of Consciousness Studies*, vol. 15, no. 1, pp. 109-128.

Rowlands, M. (2001) *The Nature of Consciousness*. Cambridge, UK: Cambridge University Press.

Roy, J.-M., Petitot, J., Pachoud, B. & Varela, F. J. (1999) "Beyond the gap: An introduction to naturalizing phenomenology", en J. Petitot, F. J. Varela, B. Pachoud & J-M Roy (eds.), *Naturalizing Phenomenology*, Stanford, CA: Stanford University Press., pp. 1-80.

Rozemond, M. (2006) "The nature of the mind", en S. Gaukroger (ed.), *The Blackwell Guide to Descartes' Meditations*, Oxford, UK: Blackwell, pp. 48-66.

Rubia, F. J. (2000) *El cerebro nos engaña*. Madrid: Temas de Hoy.

Rudder Baker, L. (1999) "Folk psychology", en R. A.Wilson & F. C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA: MIT Press, pp. 319-320.

Ruiz-Vargas, J. M. & López-Frutos, J. M. (2014) "Memoria", en P. Enríquez de Valenzuela (coord.), *Neurociencia cognitiva*, Madrid: Sanz y Torres, pp. 195-230.

Rumelhart, D. E., McClelland, J. L. & the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1*. Cambridge, MA: MIT Press.

Russel Wallace, A. (1870) *Contributions to the Theory of Natural Selection*. London: Macmillan. [Reimpreso en D. Knight (ed.), Volumen IX, pte. 1ª de *The Evolution Debate 1813-1870*, New York: Routledge, 2003].

Russell, B. A. W. (1945) *History of Western Philosophy*. New York: Simon & Schuster. [Trad. de J. Gómez de la Serna & A. Dorta, Madrid: Espasa/RBA, 2009, con prólogo de Jesús Mosterín].

Ruvinsky, A. (2010) *Genetics and Randomness*. Boca Raton, FL: CRC.

Ryle, G. (1949) *The Concept of Mind*. New York: Barnes and Noble [Ed. actual, con introducción de D. C. Dennett, Chicago: University of Chicago Press, 2007]. [Trad. de E. Rabossi, Barcelona: Paidós, 2005].

Ryle, G. (1970) “Autobiographical”, en O. P. Wood & G. Pitcher (eds.), *Ryle*, London: Macmillan, pp. 1-15.

Ryle, G. (1971) *Critical Essays: Collected Papers Volume 1*. New York: Routledge (2009).

Sáez Rueda, L. (2002) *El conflicto entre continentales y analíticos. Dos tradiciones filosóficas*. Barcelona: Crítica.

Salmon, W. C. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Salmon, W. C. (1999) *Causality and Explanation*. New York: Oxford University Press.

Savitt, S. (1974) “Rorty's disappearance theory”, *Philosophical Studies*, vol. 28, pp. 433-436.

Schnakers, C., Perrin, F., Schabus, M., Hustinx, R., Majerus, S., Moonen, G., et al. (2009) “Detecting consciousness in a total locked-in ayndrome: An active event related paradigm”, *Neurocase*, vol. 25, no. 4, pp. 271-277.

Schank, R. C. (1999) *Dynamic Memory Revisited*. Cambridge, UK: Cambridge University Press.

Schank, R. C. & Abelson, R. P. (1977) *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, R. C. & Riesbeck, C. K. (eds.) (1981) *Inside Computer Understanding: Five Programs Plus Miniatures*. New Haven, CT: Lawrence Erlbaum Associates.

Schank, R. C. & Yale A. I. Project (1975) “SAM: A story understander”, *Research Report 43*, Department of Computer Science, New Haven, CON: Yale University Press.

Schooler, J. W. (2002), “Re-representing consciousness: dissociations between experience and metaconsciousness”, *Trends in Cognitive Sciences*, vol. 6, no. 8 pp. 339-344.

Schwartz, S. P. (2012) *A Brief History of Analytic Philosophy: From Russell to Rawls*. Chichester, UK: Wiley-Blackwell.

Schwitzgebel, E. (2011) *Perplexities of Consciousness*. Cambridge, MA: MIT Press.

Scott, A. (1995) *Stairway to the Mind: The Controversial New Science of Consciousness*. New York: Copernicus, Springer.

Seager, W. E. (1999) *Theories of Consciousness. An Introduction and Assessment.* London: Routledge.

Seager, W. E. (1993) "Fodor's theory of content: Problems and objections", *Phiosophy of Science*, vol. 60, no. 2, pp. 262-277.

Searle, J. R. (1964) "How to derive 'Ought' from 'Is'", *Philosophical Review*, vol. 73, no. 1, pp. 43-58.

Searle, J. R. (1969) *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press. [Trad. de L. M. Valdés Villanueva, Madrid: Cátedra, 1990].

Searle, J. R (1975) "Indirect speech acts", en P. Cole & J. L. Morgan (eds.), *Syntax and Semantics Volume 3: Speech Acts*, New York: Academic Press, pp. 59-82.

Searle, J. R. (1978) "Literal meaning", *Erkenntnis*, vol. 13, no. 1, pp. 207-224. [Reimpreso en Searle (1979), pp. 117-136].

Searle, J. R. (1979) *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge, UK: Cambridge University Press.

Searle, J. R. (1980) "Minds, brains, and programs", *The Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417-457. [Reimpreso en J. Haugeland (ed.), *Mind Design II*. Cambridge, MA: MIT Press, 1997, pp. 183-204].

Searle, J. R. (1981a) "Intentionality and method", *The Journal of Philosophy*, vol. 78, no. 11, pp. 720-733.

Searle, J. R. (1981b) "Analytic philosophy and mental phenomena", *Midwest Studies in Philosophy*, vol. 6, no. 1, pp. 405-423.

Searle, J. R. (1983) *Intentionality. An Essay in the Philosophy of Mind*. Cambridge, UK: Cambridge University Press. [Trad. de U. Benítez & L. M. Valdés Villanueva, Madrid: Tecnos, 1992].

Searle, J. R. (1984a) "Intentionality and its place in nature", *Synthese*, vol. 61, no. 1, pp. 3-16.

Searle, J. R. (1984b) *Minds, Brains, and Science. The 1984 Reith Lectures*. Cambridge, MA: Harvard University Press. [Trad. de L. M. Valdés Villanueva, Madrid: Cátedra, 1985].

Searle, J. R. (1984c) "Minds and brains without programs", texto procedente de una lección impartida durante las *Reith Lectures* en 1984 e impreso por vez primera en C. Blakemore & S. A. Greenfield (eds.), *Mindwaves: Thoughts on Intelligence, Identity, and Consciousness*, Oxford: Blackwell, 1987, pp. 208-233.

Searle, J. R. (1989a) "Consciousness, unconsciousness and intentionality", *Philosophical Topics*, vol. 17, no. 1, pp. 193-209.

Searle, J. R. (1989b) "Reply to Jacquette", *Philosophy and Phenomenological Research*, vol. 49, no. 4, pp. 701-708.

Searle, J. R. (1990a) "Collective intentions and actions", en P. Cohen, J. Morgan & M. E. Pollack (eds.), *Intentions in Communications*, Cambridge, MA: MIT Press, pp. 401-416.

Searle, J. R. (1990b) "Consciousness, explanatory inversion and cognitive science", *Behavioral and Brain Sciences*, vol. 13, no. 4, pp. 585-596.

Searle, J. R. (1990c) "Who is computing with the brain?", *Behavioral and Brain Sciences*, vol. 13, no. 4, pp. 632-642.

Searle, J. R. (1990d) "Is the brain's mind a computer program?", *Scientific American*, vol. 262, no. 1, pp. 20-25 (en algunas ediciones 26-31).

Searle, J. R. (1990e) "Is the brain a digital computer?", *Proceedings and Addresses of the American Philosophical Association*, vol. 64, no. 3, pp. 21-37.

Searle J. R. (1991a) "Response: Perception and the satisfaction of intentionality", en E. Lepore & R. Van Gulick (eds.), *John Searle and his Critics*, Oxford: Blackwell, pp. 181-192.

Searle, J. R. (1991b) "The background of intentionality and action", en E. Lepore & R. Van Gulick (eds.), *John Searle and his Critics*, Oxford: Blackwell, pp. 289-299.

Searle, J. R. (1992) *The Rediscovery of the Mind*. Cambridge, MA: MIT Press. [Trad. de L. M. Valdés Villanueva, Barcelona: Crítica, 1996].

Searle, J. R. (1993a) "The problem of consciousness", en J. R. Searle, *Consciousness and Language*, Cambridge, UK: Cambridge University Press, 2002, pp. 7-17.

Searle, J. R. (1993b) "The failures of computationalism", *Think*, vol. 2 no. 1, pp. 68-73.

Searle, J. R. (1994a) "Animal minds", en J. R. Searle, *Consciousness and Language*, Cambridge, UK: Cambridge University Press, 2002, pp. 61-76.

Searle, J. R. (1994b) "Searle, John, R." en S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Cambridge, MA: Blackwell, pp. 544-550.

Searle, J. R. (1995a) *The Construction of Social Reality*. New York: The Free Press. [Trad. de Antoni Domènech, Barcelona: Paidós, 1997].

Searle, J. R. (1995b) "Consciousness, the brain and the connection principle: A reply", *Philosophy and Phenomenological Research*, vol. 55, no.1, pp. 217-232.

Searle, J. R. (1997a) *The Mystery of Consciousness*. New York: The New York Review of Books. [Trad. de A. Domènech Figueras, Barcelona: Paidós, 2000].

Searle, J. R. (1997b) *Mind, Language and Society: Philosophy in the Real World*. New York: Basic Books. [Trad. de J. Alborés Rey, Madrid: Alianza, 2001].

Searle, J. R. (1997c) "The explanation of cognition", *Royal Institute of Philosophy Supplement*, vol. 42, pp. 103-126.

Searle, J. R. (1998) "How to study consciousness scientifically", *Brain Research Reviews*, Special Issue, vol. 26, no 3, pp. 379-387.

Searle, J. R. (1999a) "Philosophy and the habits of critical thinking: Conversations with John R. Searle", H. Kreisler, *Conversations with History*. UC Berkeley: Institute of International Studies.
URL: <http://globetrotter.berkeley.edu/people/Searle/searle-con0.html>.

Searle J. R. (1999b) "The future of philosophy", *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, vol. 354, no. 1392, pp. 2069-2080. [Versión española de J. I. Guarino aparecida en *Eikasia. Revista de Filosofía*, no. 32, mayo de 2010, pp. 1-32].

Searle, J. R. (1999c) "El trasfondo de la intencionalidad", *Teorema*, vol. 18, no. 1, pp. 7-18.

Searle, J. R. (2000a) "Consciousness", *Annual Review of Neuroscience*, vol. 23, no. 1, pp. 557-578. [Reimpreso en J. R. Searle, *Consciousness and Language*, Cambridge, UK: Cambridge University Press, 2002, pp. 36-60].

Searle, J. R. (2000b) "Mental causation, conscious and unconscious: A reply to Anthonie Meijers", *International Journal of Philosophical Studies*, vol. 8, no. 2, pp. 171-177.

Searle, J. R. (2001) *Rationality in Action*. Cambridge, MA: MIT Press.

Searle, J. R. (2002a) *Consciousness and Language*. Cambridge, UK: Cambridge University Press.

Searle, J. R. (2002b) "Why I am not a property dualist?", *Journal of Consciousness Studies*, vol. 9, no. 12, pp. 57-64.

Searle, J. R. (2003) "Reply to Jaegwon Kim on mental causation". URL: <http://ist-socrates.berkeley.edu/~jsearle/132/>.

Searle, J. R. (2004a) *Mind: A Brief Introduction*. New York: Oxford University Press.

Searle, J. R. (2004b) *Freedom & Neurobiology. Reflections on Free Will, Language and Political Power*. New York: Columbia University Press. [Trad. de M. Candel Barcelona: Paidós, 2005].

Searle, J. R. (2005) "The phenomenological illusion", en M. E. Reicher, J. C. Marek (eds.), *Erfahrung und Analyse. Experience and Analysis*, Wien: 27th International Wittgenstein Symposium, pp. 317-336.

Searle, J. R. (2006) "What is to be done", *Topoi*, vol. 25, no. 1, pp. 101-108.

Searle, J. R. (2007a) "Biological naturalism", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 325-334.

Searle, J. R. (2007b) "Dualism revisited", *Journal of Physiology-Paris*, vol. 101, no. 4-6, pp. 169-178.

Searle, J. R. (2007c) "Putting consciousness back in the brain: Reply to Bennett and Hacker", en M. R. Bennett, D. C. Dennett, P. M. S. Hacker & J. R. Searle, *Neuroscience & Philosophy. Brain, Mind, & Languge*, New York: Columbia University Press, pp. 97-126. [Trad. de R. Filella, Barcelona: Paidós, 2008].

Searle, J. R. (2009) "Biological naturalism", en T. Bayne, A. Cleeremans & P. Wilken (eds.), *The Oxford Companion to Consciousness*, Oxford, UK: Oxford University Press, pp. 107-109.

Sellars, R. W. (1927) "Why naturalism and not materialism", *Philosophical Review*, vol. 36, pp. 216-225. [Reimpreso en. W. Preston Warren (ed.), *Principles of Emergent Realism: Philosophical Essays of Roy Wood Sellars*, St. Louis, MO: Warren H. Green, 1970, pp. 131-139].

Sellars W. S. (1956) "Empiricism and the philosophy of mind", en Feigl, H. & Scriven, M. (eds.), *The Foundations of Science and the Concepts of of Psychology and Psychoanalysis: Minnesota Studies in the Philosophy of Science, Vol. 1*, Minneapolis: University of Minnesota Press, pp. 253-329.

Sellars, W. S. (1962) "Philosophy and the scientific image of man", en R. Colodny (ed.), *Frontiers of Science and Philosophy*, Pittsburgh: University of Pittsburgh Press, pp. 35-78. [Citamos la reimpresión publicada en Sellars (1963), pp. 1-40].

Sellars, W. S. (1963) *Empiricism and the Philosophy of Mind*. London: Routledge & Kegan Paul Ltd.

Seth, A. K. & Edelman, G. M. (2009) "Consciousness and complexity", en B. Meyer (ed.), *Encyclopedia of Complexity and Systems Science*, Vol. 2, Berlin: Springer, pp. 1424-1443.

Shallice, T. (1972) "Dual functions of consciousness", *Psychological Review*, vol. 79, no. 5, pp. 383-393.

Shani, I. (2007) "Consciousness and the first person. A critical appraisal of Searle's connection principle", *Journal of Consciousness Studies*, vol. 14, no. 12, pp. 57-91.

Shani, I. (2008) "Against consciousness chauvinism", *Monist*, vol. 91, no. 2, pp. 294-323.

Shannon, C. E. (1948) "A mathematical theory of communication", *Bell System Technical Journal*, vol. 27, pp. 379-423/623-656.

Shapin, S. (1994) *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Chicago: University of Chicago Press.

Shapin, S. (1996) *The Scientific Revolution*. Chicago: University of Chicago Press. [Trad. de J. Romo Feito, Barcelona: Paidós, 2000].

Sheehy, N. (2004) *Fifty Key Thinkers in Psychology*. London: Routledge.

Sheets-Johnstone, M. (1998) "Consciousness: A natural history", *Journal of Consciousness Studies*, vol. 5, no. 3, pp. 260-294.

Shepherd, G. (1972) "The neuron doctrine: A revision of functional concepts", *Yale Journal of Biology and Medicine*, vol. 45, pp. 584-599.

Shewmon, D. A. (1997) "Recovery from 'brain death': A neurologist's apologia", *Linacre Quarterly*, vol. 64, no. 1, pp. 30-96.

Shoemaker, S. (1994a) "Phenomenal character", *Noûs*, vol. 28, pp. 21-38.

Shoemaker, S. (1994b) "Self-knowledge and 'inner sense'. Lecture III: The phenomenal character of experience", *Philosophy and Phenomenological Research*, vol. 54, no. 2, pp. 291-314.

Sider, T. (1996) "Intrinsic properties", *Philosophical Studies*, vol. 83, no. 1, pp. 1-27.

Sierra, J. C., Luna Villegas, G., Fernández Guardiola, A. & Buela Casal, G. (1993) "Evaluación de la activación y de la vigilancia", *Revista Latinoamericana de Psicología*, vol. 25, no. 3, pp. 433-452.

Siewert, C. (1999) *The Significance of Consciousness*. Princeton, New Jersey: Princeton University Press.

Sigüenza, J. A. (1993) *Neurocomputación. Cómo funciona el cerebro*. Madrid: Eudema.

Sloman, A. & Chrisley, R. L. (2003) "Virtual machines and consciousness", *Journal of Consciousness Studies*, vol. 10, nos. 4-5, pp. 133-172.

Smith, A. P. (2008) *The Dimensions of Experience: A Natural History of Consciousness*. Bloomington, IN: Xlibris.

Smith, S. M., Brown, H. O., Toman, J. E. P. & Goodman, L. S. (1947) "The lack of cerebral effects of d-tubocurarine", *Anesthesiology*, vol. 8, no. 1, pp. 1-14.

Soddu, A., Boly1, M., Nir, Y., Noirhomme, Q., Vanhaudenhuyse, A., Demertzi, A., et al. (2009) "Reaching across the abyss: Recent advances in functional magnetic resonance imaging and their potential relevance to disorders of consciousness", *Progress in Brain Research*, vol. 177, pp. 261-274.

Solms, M. & Turnbull, O. (2002) *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*. New York: Other Press. [Trad. de D. Jaramillo, prólogo de Oliver W. Sacks, Bogotá: Fondo de Cultura Económica, 2004].

Sorem, E. (2010) "Searle, materialism, and the mind-body problem", *Perspectives: International Postgraduate Journal of Philosophy*, vol. 3, no.1, pp. 30-54.

Sorger, B., Dahmen, B., Reithler, J., Gosseries, O., Maudoux, A., Laureys, S. & Goebel, R. (2009) "Another kind of 'BOLD response': Answering multiple-choice questions via online decoded single-trial brain signals", *Progress in Brain Research*, vol. 177, pp. 275-292.

Sperry, R. W. (1984) "Consciousness, personal identity and the divided brain", *Neuropsychologia*, vol., 22, no. 6, pp. 611-673.

Squire, L. R., Bloom, F. E., Spitzer, N. C., du Lac, S., Gosh, A. & Berg, D. (eds.) (2008) *Fundamental Neuroscience* (3rd ed.). San Diego, CA: Academic Press.

Stamp Dawkins, M. S. (2000) "Animal minds and animal emotions", *Integrative and Comparative Biology (American Zoologist)*, vol. 40, no. 6, pp. 883-888.

Stapp, H. P. (2007) "Quantum mechanical theories of consciousness", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 300-312.

Stapp, H. P. (2007/2011) *Mindful Universe: Quantum Mechanics and the Participating Observer* (2nd ed.). Berlin: Springer.

Stapp, H. P. (2011) "Searle's 'Dualism revisited'", *Review of Contemporary Philosophy*, vol. 10, pp. 141-149.

Stich, S. P. (1987) "Review of Searle, J., 'Minds, brains and science'", *Philosophical Review*, vol. 96, no. 1, pp. 129-133.

Stigol, N. (2001) "Representacionalismo y qualia", *Teorema*, vol. 20, no. 3, pp. 31-39.

Stoljar, D (2004) "The argument from diaphanousness", en M. Ezcurdia, R. Stainton & C. Viger (eds.), *New Essays in the Philosophy of Language and Mind: Canadian*

*Journal of Philosophy Supplementary Volume*, Calgary: University of Calgary Press, pp. 341-390.

Stoljar, D. (2005) "Physicalism and phenomenal concepts", *Mind and Language*, vol. 20, no. 5, pp. 469-494.

Stout, G. F. (1899) *A Manual of Psychology*. New York: Hinds, Noble & Eldredge Publishers. [Ed. actual: Washington, DC: University Publications of America, 1977]

Stout, G. F. (1931) *Mind and Matter*. Cambridge, UK: Cambridge University Press. [Ed. actual: Cambridge, UK: Cambridge University Press, 2011].

Strawson, G. (2005) "Intentionality and experience: Terminological preliminaries", en D. W. Smith & A. L. Thomasson (eds.), *Phenomenology and Philosophy of Mind*, Oxford, UK: Oxford University Press, 41-66.

Strawson, G. (2006) "Realistic monism: Why physicalism entails panpsychism", *Journal of Consciousness Studies*, vol. 13, nos. 10-11, pp. 3-31. [Reimpreso en A. Freeman (ed.), *Consciousness and its Place in Nature: Does Physicalism entail Panpsychism?*, Exeter: Imprint Academic, 2006, pp. 3-31].

Strawson, G. (2008) *Real Materialism and Other Essays*. Oxford, UK: Oxford University Press.

Strawson, G. (1994/2010) *Mental Reality* (2nd ed.). Cambridge, MA: MIT Press, Bradford Books [Trad. de M. A. Galmarini, Barcelona: Prensa Ibérica, 1997].

Strawson, G. (2010) "Radical self-awareness", en M. Siderits, E. Thompson & D. Zahavi (eds.), *Self, No Self? Perspectives From Analytical, Phenomenological, and Indian Traditions*, New York: Oxford University Press, pp. 274-307.

Strawson, P. F. (1950) "On referring", *Mind*, vol. 59, no. 235, pp. 320-344. [Reimpreso en R. R. Ammerman (ed.), *Classics of Analytic Philosophy*, Indianapolis, IN: Hackett Publishing, 1990, pp. 315-334].

Strawson, P. F. (1952) *Introduction to Logical Theory*. London: Metheum, 1990.

Strawson, P. F. (1985) *Scepticism and Naturalism: Some Varieties*. New York: Columbia University Press.

Stroud, B. (1991) "The background of thought", en E. Lepore & R. Van Gulick (eds.), *John Searle and his Critics*, Oxford: Blackwell, pp. 245- 258.

Stroud, B. (1996) "The charm of naturalism", *Proceedings and Addresses of the American Philosophical Association*, vol. 70, no. 2, pp. 43-55.

Sturgeon, S. (1994) "The epistemic view of subjectivity", *Journal of Philosophy*, vol. 91, no. 5, pp. 221-235.

Sturgeon, S. (2000) *Matters of Mind: Consciousness, Reason and Nature*. London: Routledge.

Styles, E. (2006) *The Psychology of Attention*. Hove: Psychology Press. [Trad. de A. Atienza Díez, Madrid: Editorial Universitaria Ramón Areces, 2010.]

Sutherland, N. S. (ed.) (1989) *The International Dictionary of Psychology*. New York: Continuum.

Swanson, L. W. (2012) *Brain Architecture: Understanding the Basic Plan* (2nd ed.). New York: Oxford University Press.

Swinburne, R. (1986) *The Evolution of the Soul*. Oxford, Clarendon.

Swinburne, R. (2013) *Mind, Brain, and Free Will*. Oxford, UK: Oxford University Press.

Tarnas, R. (1991) *The Passion of the Western Mind: Understanding the Ideas That Have Shaped Our World View*. Ballantine Books. [Trad. de M. A. Galmarini, Vilaür: Atalanta, 2008].

Taylor, J. G. (1999) *The Race for Consciousness*. Cambridge, MA: MIT Press.

Tandon, P. N. (2000) "The decade of the brain: A brief review", *Neurology India*, vol. 48, no. 3, pp. 199-207.

Thagard, P. (2005) *Mind: Introduction to Cognitive Science* (2nd ed.). Cambridge, MA: MIT Press. [Trad. de J. Barba & S. Jawerbaum, Buenos Aires: Katz, 2008].

Thagard, P. (2007) "Introduction to the philosophy of psychology and cognitive science", en P. Thagard (volume ed.), *Handbook of the Philosophy of Science. Philosophy of Psychology and Cognitive Science*, Amsterdam: North-Holland Elsevier, pp. ix-xvii.

Thau, M. (2002) *Consciousness and Cognition*. New York: Oxford University Press.

Thomasson, A. L. (2002) "Phenomenology and the development of analytic philosophy", *Southern Journal of Philosophy*, vol. 40 (suppl.), pp. 115-142.

Thompson, E. (2007) *Mind in Life. Biology, Phenomenology, and the Sciences of Mind*, Cambridge, MA: Harvard University Press.

Thompson E., Lutz, A. and Cosmelli, D. (2005) "Neurophenomenology: An Introduction for Neurophilosophers", en A. Brook & K. Akins (eds.), *Cognition and the Brain: The Philosophy and Neuroscience Movement*, New York: Cambridge University Press, pp. 40-97.

Thompson, E., Noë, A. & Pessoa, L. (1999) "Perceptual completion: A case study in phenomenology and cognitive science", en J. Petitot, F. J. Varela, B. Pachoud & J-M Roy (eds.), *Naturalizing Phenomenology*, Stanford, CA: Stanford University Press., pp. 161-195.

Thompson, E. & Varela, F. J. (2001) "Radical embodiment: Neural dynamics and consciousness", *Trends in Cognitive Sciences*, vol. 5, no. 10, pp. 418-425.

Tinbergen, N. (1963) "On aims and methods of ethology", *Zeitschrift für Tierpsychologie*, vol. 20, no. 4, pp. 410-433.

Toates, F. M. (1998) "The interaction of cognitive and stimulus-response processes in the control of behaviour", *Neuroscience and Biobehavioural Reviews*, vol. 22, no. 1, pp. 59-83.

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946) "Studies in spatial learning. Place learning versus response learning", *Journal of Experimental Psychology*, vol. 36, no. 3, pp. 221-229.

Tononi, G. (2004) "An information integration theory of consciousness", *BMC Neuroscience*, vol. 5, art. no. 42. DOI: <10.1186/1471-2202-5-42>.

Tononi, G. (2008) "Consciousness as integrated information: A provisional manifesto", *The Biological Bulletin*, vol. 215, no. 3, pp. 216-242.

Tononi, G. (2012a) *Phi: A Voyage from the Brain to the Soul*. New York: Pantheon Books.

Tononi, G. (2012b) "Integrated information theory of consciousness: An updated account", *Archives Italiennes de Biologie*, vol. 150, no. 2-3, pp. 290-326.

Tononi, G. & Edelman, G. M. (1998) "Consciousness and complexity", *Science*, vol. 282, no. 5395, pp. 1846–1851.

Torres Bares, C. & Escarabajal Arrieta, M. D. (2005) "Psicofarmacología: Una aproximación histórica", *Anales de Psicología*, vol. 21, no. 2, pp. 199-212.

Tulving, E. (1972) "Episodic and semantic memory", en E. Tulving & W. Donaldson (eds.), *Organization of Memory*, New York: Academic Press, pp. 382-402.

Tulving, E. (2000) "Introduction (to VI, 'Memory')", en M. S. Gazzaniga (ed.), *The New Cognitive Neurosciences*, (2nd ed.), Cambridge, MA: MIT Press, pp. 727-732.

Turing, A. M. (1936). "On computable numbers, with an application to the Entscheindungproblem", *Proceedings of the London Mathematical Society*, ser. 2, vol. 42, pp. 230-265. [Reimpreso en Davis, M. (ed.), *The Undecidable: Basic*

*Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press, 1965, pp. 115-154].

Turing, A. M. (1950) "Computing machinery and intelligence", *Mind*, vol. 59, no. 236, pp. 433-460. [Reimpreso Boden, M. A. (ed.), *The Philosophy of Artificial Intelligence*. Oxford, UK: Oxford University Press, 1990, 40-66].

Tye, M. (1992) "Visual qualia and visual content", en T. Crane (ed.), *The Contents of Experience*, Cambridge, UK: Cambridge University Press, pp. 158-176.

Tye, M. (1995) *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Tye, M. (2000) *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

Tye, M. (2002) "Representationalism and the transparency of experience", *Noûs*, vol. 36, no. 1, pp. 137-151.

Tye, M. (2007) "Philosophical problems of consciuousness", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 23-35.

Tye, M. (2008) "The experience of emotions. An intentionalist theory", *Revue Internationale de Philosophie*, vol. 62, no. 243, pp. 25-50.

Tye, M. (2009) *Consciousness Revisted: Materialism without Phenomenal Concepts*. Cambridge, MA: MIT Press.

Tyndall, J. (1871) *Fragments of Science for Unscientific People: A Series of Detached Essays, Lectures, and Reviews*. London: Longmans Green and Co. [Ed. actual: London: Forgotten Books, 2013].

Valdés, L. M. (2009) "John Searle: Intencionalidad, conciencia y libre albedrío", en L. M. Valdés (coord.) "Nuevas Variedades del pensamiento analítico", cap. 19 (pp. 529-595) de M. Garrido, L. M. Valdés & L. Arenas (coords.), *El legado filosófico y científico del siglo XX* (3ª ed.), Madrid: Cátedra, pp. 577-581.

Valenstein, E. (2005) *The War of the Soups and the Sparks. The Discovery of Neurotransmitters and the Dispute over How Nerves Communicate*. New York: Columbia University Press.

Vallentine, P. (1997) "Intrinsic properties defined", *Philosophical Studies*, vol. 88, no. 2, pp. 209-219.

Van Gelder, T. (1995) "What might cognition be, if not computation?", *The Journal of Philosophy*, vol. 92, no. 7, pp. 345-381.

Van Gelder, T. (1999a) "Dynamic approaches to cognition", en R. A.Wilson & F. C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA: MIT Press, pp. 243-245.

Van Gelder, T. (1999b) "Wooden iron? Husserlian phenomenology meets cognitive science", en J. Petitot, F. J. Varela, B. Pachoud & J-M. Roy (eds.), *Naturalizing Phenomenology*, Stanford, CA: Stanford University Press, pp. 245-265.

Van Gulick, R. (2001) "Reduction, emergence and other recent options on the mind/body problem. A philosophic overview", vol. 8, no. 9-10, 2001 , pp. 1-34

Van Gulick, R. (2011) "Consciousness", en E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition)*.
URL: <http://plato.stanford.edu/archives/sum2011/entries/consciousness/>.

Varela, F. J. (1995) "Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony", *Biological Research*, vol. 28, no. 1, pp. 81-95.

Varela, F. J. (1996) "Neurophenomenology: A methodological remedy for the hard problem", *Journal of Consciousness Studies*, vol. 3, no. 4, pp. 330-349.

Varela, F. J. (1997) "The naturalization of phenomenology as the transcendence of nature: Searching for generative mutual constraints", *Alter: Revue de Phénoménologie*, vol. 5, no. 4, pp. 355-385.

Varela, F. (2001) "Cerebro y concienca", en J. Aguado Sobrino, F. Mora Teruel & J. M. Segovia de Arana (coords.), *Ciencia y sociedad: La tercera cultura*, Oviedo: Nobel, pp. 167-194. [Reimpreso en F. Mora (coord.) *Esplendores y miserias del cerebro*, Madrid: Fundación Santander Central Hispano, 2004, pp. 226-257].

Varela, F. J. (1999) "The specious present: a neurophenomenology of time consciousness", en J. Petitot, F. J. Varela, B. Pachoud & J-M Roy (eds.), *Naturalizing Phenomenology*, Stanford, CA: Stanford University Press., pp. 266-314.

Varela, F. J., Thompson, E. & Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press [Trad. de C. Gardini, Barcelona: Gedisa, 1992].

Varela, F. J. & Shear, J. (1999) "First-person methodologies: What, why, how?", *Journal of Consciousness Studies*, vol. 6, nos. 2-3, pp. 1-14.

Varela, F. J., Lachaux, J-P., Rodriguez, E. & Martinerie, J. (2001) "The brainweb: phase synchronization and large-scale integration", *Nature Reviews Neuroscience*, vol. 2, no. 4, pp. 229-239.

Varela, F. J. & Thompson, E. (2003) "Neural synchrony and the unity of mind: A neurophenomenological perspective", en A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, and Dissociation*, New York: Oxford University Press, pp. 266-287.

Velmans, M. (2009) *Understanding Consciousness* (2nd ed.). New York: Routledge.

Velmans, M. (1996) "An introduction to the science of consciousness", en M. Velmans (ed.), *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews*, London: Routledge, pp. 1-22.

Vohra, A. (1989) *Wittgenstein's Philosophy of Mind*. La Salle, IL.: Open Court.

Vicari, G. (2008) *Beyond Conceptual Dualism: Ontology of Consciousness, Mental Causation, and Holism in John R. Searle's Philosophy of Mind*. Amsterdam: Rodopi B. V.

Vicente, A. (2001) "El principio del cierre causal del mundo físico", *Crítica. Revista Hispanoamericana de Filosofía*, vol. 33, no. 99, pp. 3-17.

Vilarroya, O. (2002a) *La disolución de la mente: Una hipótesis sobre cómo siente, piensa y se comunica el cerebro*. Barcelona: Tusquets. [Ed. inglesa, Amsterdam: Rodopi, 2002].

Vilarroya, O. (2002b) *Paraula de robot: intelligència artificial i comunicació*. Alzira: Bromera; València: Publicacions de la Universitat de València. [Trd. de L. Valencia, Valencia: Publicacions de la Universitat de València, 2006].

Villanueva, E. (1995) "Conciencia", en F. Broncano (ed.), *La mente humana*, Madrid: Trotta, pp. 385-400.

Vincent, J-D. (2007) *Voyage extraordinaire au centre du cerveau*. Paris: Odile Jacob. [Trad. de C. Zelich Martínez, Barcelona: Anagrama, 2009].

Viñuela Fernández, F. (2007) "Trastornos de las funciones visuoespacial y constructiva", en J. Peña-Casanova (coord.), *Neurología de la conducta y neuropsicología*, Buenos Aires/Madrid: Médica Panamericana, pp. 233-242.

Vitiello, G. (2003) "Quantum dissipation and information: A route to consciousness modelling", *NeuroQuantology*, vol. 1 no. 2, pp. 266-279.

Von Neumann, J. (1958) *The Computer and the Brain*. New Haven/London: Yale Univesity Press. [Ed. actual, con prólogos de Ray Kurzweil y los Churchland y prefacio de Klara von Neumann, New Haven, CON: Yale University Press, 2012. Trad., con introducción de P. R. Halmos y prefacio de Klara von Neumann, de J. Borrell, Barcelona: Antoni Bosch, 1980].

Von Weizsäcker, C. F. F. (1985/2006) *Aufbau der Physik*. Munich: Hanser. [Trad. inglesa de H. Biritz. Ed., revisada y ampliada, de T. Görnitz & H. Lyre, Dordrecht: Springer, 2006].

Walling, P. T. & Hicks, K. N. (2003) "Dimensions of consciousness", *Baylor University Medical Center Proceedings*, vol. 16, no.2, pp. 162–166.

Walter, H. (1999) *Neurophilosophie der Willensfreiheit. Von libertarischen Illusionen zum Konzept natürlicher Autonomie*. Paderborn: Mentis. [Tard. inglesa de C. Klohr, Cambridge, MA: MIT Press, 2001, por donde citamos].

Wan, X. I., Ambinder, M. S. & Simons, D. J. (2009) "Change blindness", en T. Bayne, A. Cleeremans & P. Wilken (eds.), *The Oxford Companion to Consciousness*, Oxford, UK: Oxford University Press, pp. 130-133.

Ward, J. (1918) *Psychological Principles*. Cambridge, UK: Cambridge University Press. [Ed. actual, por donde citamos, Cambridge, UK: Cambridge University Press, 2013].

Ward, L. M. (2003) "Synchronous neural oscillations and cognitive processes", *Trends in Cognitive Sciences*, vol. 7, no. 12, pp. 553-559.

Wasserman, E. A. & Zentall, T. R. (2012) "Introduction", en T. R. Zentall & E. A. Wasserman (eds.), *The Oxford Handbook of Comparative Cognition*, New York: Oxford University Press, pp. 1-8.

Watson, J. B. (1913) "Psychology as the behaviorist views it", *Psychological Review*, no. 20, pp. 158-177. [Reimpreso en W. Lyons (ed.), *Modern Philosophy of Mind*, London: Everyman, 1995, pp. 24-42].

Watson, J. B. (1924) *Behaviourism*. New York: People's Institute. [Sucesivas reimpresiones en New York: Norton. Ed. revisada de 1930 citada en esta tesis: New Jersey: Transaction, 1998. Trad. de O. Poli, Buenos Aires: Paidós, 1972].

Watt, D. F. (1998) "The implications of affective neuroscience for extended reticular thalamic activating system theories of consciousness", *Online Electronic Seminar for the Association for the Scientific Study of Consciousness*. URL: <http:// server.phil.vt.edu/assc/esem.html>.

Watt, D. F. (1999) "At the intersection of emotion and consciousness: Affective neuroscience and extended reticular thalamic activating system (ERTAS) theories of consciousness", en S. R. Hameroff, A. W. Kaszniak & D. J. Chalmers (eds.), *Toward a Science of Consciosness III: The Third Tucson Discussions and Debates*, Cambridge, MA: MIT Press, pp. 215-229.

Waugh, N. C. & Norman, D. A. (1965) "Primary memory", *Psychological Review*, vol. 72, no. 2, pp. 89-104.

Weiskrantz, L. (2007) "The case of blindsight", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 175-180.

White, S. L. (1986) "Curse of the qualia", *Synthese*, vol. 68, no. 2, pp. 333-368.

Whitehead, A. N. (1929) *Process and Reality: an Essay in Cosmology*. New York : Macmillan. [Ed. actual de D. R. Griffin & D. W. Sherburne, New York: Free Press, 1978. Trad. de J. Rovira Armengol., Buenos Aires: Losada, 1956].

Whitehead, A. N. (1933) *Adventures of Ideas*. New York: Macmillan.

Whorf, B. L. (1956) *Thought, Language, and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press. [Primera ed. de J. B. Carroll. Ed. actual de J. B. Carroll, S. C. Levinson & P. Lee, Cambridge, MA: MIT Press, 2012].

Wider, K. V. (1997) *The Bodily Nature of Consciousness: Sartre and Contemporary Philosophy of Mind*. Ithaca, NY: Cornell University Press.

Wilkes, K. V. (1984) "Is consciousness important?", *British Journal for the Philosophy of Science*, vol. 35, no. 3, pp. 223-243.

Wilkes, K. V. (1988) "Yishi, duh, um and consciousness", en A. J. Marcel & E. Bisiach (eds), *Consciousness in Contemporary Science*, Oxford, UK: Oxford University Press, pp. 16-41.

Wilkes, K. V. (1995) "Losing consciousness", en T. Metzinger (ed.), *Conscious Experience*, Paderborn: Ferdinand Schoningh, pp. 97-106.

Williams, P. (1998) *The Reflexive Nature of Awareness*. London: Curzon Press.

Williamson, T. (2011) "What is Naturalism?", *The New York Times*, Sep. 4. URL: <http://opinionator.blogs.nytimes.com/2011/09/04/what-is-naturalism/>.

Wilson, R. A. & Keil, F. C. (eds.) (1999) *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.

Wittgenstein, L. J. J. (1921) *Logisch-Philosophische Abhandlung*. London: Routledge & Kegan Paul, 1922. [Ed. inglesa de C. K. Ogden, trad. del original publicado en *Annalen der Naturphilosophie*, vol. 14, nos. 3-4, pp. 185-262 de C. K. Ogden & F. P. Ramsey. Trad. castellana de J. Muñoz & I. Reguera, Madrid: Alianza, 2003].

Wittgenstein, L. J. J. (1953) *Philosophische Untersuchungen*. Oxford: Basil Blackwell. [Ed. inglesa de G. E. M. Anscombe & R. Rhees, trad. de G. E. M. Anscombe. Trad. castellana de A. García Suárez & U. Moulines, Barcelona: Crítica, 1988].

Wittgenstein, L. J. J. (1969) *Über Gewißheit*. Oxford: Basil Blackwell. [Ed. inglesa de G. E. M. Anscombe & G. H. von Wright, trad. de D. Paul & G. E. M. Anscombe. Trad. castellana de J. L. Prados & V. Raga, Barcelona: Gedisa, 1988].

Workman, L. (2014) *Charles Darwin: Shaper of Evolutionary Thinking*. Basingstoke: Palgrave Macmillan.

Yablo, S. (1999) "Intrinsicness", *Philosophical Topics*, vol. 26, nos. 1-2, pp. 479-505.

Zahavi, D. (1999) *Self-Awareness and Alterity*. Evanston, IL: Northwestern University Press.

Zahavi, D. (2002) "First-person thoughts and embodied self-awareness: Some reflections on the relation between recent analytic philosophy and phenomenology", *Phenomenology and the Cognitive Sciences*, vol. 1, no. 1, pp. 7-26.

Zangwill, O. L. (1977) "Consciousness and the brain", en R. E. Butts & J. Hintikka (eds.), *Foundational Problems in the Special Sciences. Part Two of the Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, London, Ontario, Canada, 1975*, Dordrecht: Reidel, pp. 153-165.

Zeki, S. (1993) *A Vision of the Brain*. Oxford: Blackwell. [Trad. de J. Soler, Barcelona: Ariel, 1995].

Zeki, S. (2003) "The disunity of consciousness", *Trends in Cognitive Science*, vol. 7, no. 5, pp. 214-218.

Zeki, S. (2005) "The Ferrier Lecture 1995. Behind the seen: The functional specialization of the brain in space and time", *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, vol. 360, no. 1458, pp 1145-1183.

Zeki, S. (2007) "A theory of micro-consciousness", en M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Oxford: Blackwell, pp. 580-588.

Zeki, S. & Bartels, A. (1998) "The theory of multistage integration", *Proceedings of the Royal Society, Series B: Biological Sciences*, vol. 265, no. 1412, pp. 2327-2332.

Zeki, S. & Bartels, A. (1999) "Toward a theory of visual consciousness", *Consciousness and Cognition*, vol. 8, no. 2, pp. 225-259.

Zeki, S. & Ffytche, D. (1998) "The Riddoch syndrome: Insights into the neurobiology of conscious vision", *Brain*, vol. 121, no. 1, pp. 25-45.

Zemach, E. M. (1991) “Perceptual realism, naive and otherwise”, en *John Searle and his Critics*, E. Lepore & R. Van Gulick (eds.), Oxford: Blackwell, pp 169-179.

Zeman, A. (2002) *Consciousness: A User’s Guide*. New Haven, CON: Yale University Press. [Trad. de R. Reyes-Mazzoni, México, DF: Fondo de Cultura Económica, 2009].

Zeman, A. (2008) *A Portrait of the Brain*. London: Yale University Press. [Trad. de J. Sarret Grau, Barcelona: Ediciones de Intervención Cultural, Biblioteca Buridán, 2009].