

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FILOSOFÍA

Departamento de Teoría del Conocimiento,
Estética e Historia del Pensamiento



SERES HUMANOS, YOES Y PERSONAS: UNA PROPUESTA NARRATIVA PARA EL PROBLEMA DE LA IDENTIDAD PERSONAL

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Alfonso Muñoz Corcuera

Bajo la dirección de los doctores

Mariano Rodríguez González
Ángel García Galiano

Madrid, 2013

ÍNDICE

Agradecimientos.....	5
Introducción.....	7
Capítulo 1. El Planteamiento Objetivo.....	19
1.1. El príncipe y el zapatero.....	24
1.2. Los dobles de Guy Fawkes.....	30
1.3. El señor Brownson.....	32
1.4. El científico malvado.....	41
1.5. El teletransporte.....	44
1.6. Conclusiones.....	47
Capítulo 2. El Planteamiento Narrativo Clásico.....	49
2.1. Entendiendo el giro narrativo.....	52
2.2. Los dos tesis narrativas.....	55
2.2.1. La tesis psicológica de la Narratividad.....	56
2.2.2. La tesis ética de la Narratividad.....	61
2.3. Críticas al narrativismo.....	66
2.3.1. La falsedad de las dos tesis de la Narratividad.....	69
2.3.2. El Yo fenoménico.....	73
2.4. El Nuevo Planteamiento Narrativo: problemas sin resolver.....	75
2.5. Conclusiones.....	78
Capítulo 3. Los problemas del materialismo.....	81
3.1. Lo que nos importa.....	85
3.1.1. Aproximación conceptual.....	85

3.1.2. El jefe mafioso.....	88
3.2. El materialismo frente al problema de la conciencia.....	93
3.3. El materialismo frente al problema de las personas.....	100
Capítulo 4. Una propuesta anti-realista.....	107
4.1. Críticas al realismo.....	110
4.2. Pragmatismo y el problema de la convergencia.....	115
4.3. El juego de la vida y la actitud interpretativa.....	119
4.4. Tipos de mundo.....	127
4.5. El problema de la identidad narrativa.....	135
Capítulo 5. El Nuevo Planteamiento Narrativo.....	137
5.1. El mundo de los seres humanos.....	138
5.2. El mundo de los Yoes.....	141
5.2.1. Fenomenología y narratividad.....	142
5.2.2. Supervivencia e identidad.....	148
5.3. El mundo de las personas.....	155
5.3.1. La actitud colectiva.....	155
5.3.2. La tesis social de la Narratividad.....	161
5.4. Resolviendo los experimentos mentales.....	167
Capítulo 6. Personas y personajes.....	173
6.1. Críticas recibidas.....	174
6.2. Solucionando el problema.....	179
6.2.1. Hacia un concepto apropiado de narración.....	180
6.2.2. Escepticismo y negociaciones narrativas.....	188
6.2.3. Diferencias entre personas y personajes.....	194
Conclusiones.....	201
Anexos.....	207
Anexo I. Sobre la relación de identidad.....	209
Anexo II. Introduction.....	216
Anexo III. Conclusions.....	227
Bibliografía.....	233

AGRADECIMIENTOS

Empecé mis estudios de doctorado en la Facultad de Filosofía de la Universidad Complutense de Madrid durante el curso 2008/2009, y desde entonces han sido numerosos los seres humanos, Yoes y personas que de un modo u otro han contribuido a que esté en disposición de terminarlos.

Debo agradecer en primer lugar a mis directores de tesis, el Dr. Mariano Rodríguez González y el Dr. Ángel García Galiano, por su guía y apoyo durante estos años. Sin ellos seguramente mi camino habría sido otro más difícil y menos provechoso.

También quisiera expresarles mi gratitud a los que han sido mis tutores durante las estancias de investigación que he realizado en otros centros durante este tiempo. El Dr. Galen Strawson (University of Reading, 2011), el Dr. David Chalmers (Australian National University, 2012) y el Dr. Ambrosio Velasco (Universidad Nacional Autónoma de México, 2012 y 2013). Sus comentarios sobre mi trabajo sin duda contribuyeron a mejorarlo.

Durante mi investigación partes de mi tesis fueron presentadas en numerosos congresos y seminarios a cuyos organizadores agradezco el espacio que me cedieron. Las preguntas y comentarios de los participantes enriquecieron enormemente mi perspectiva sobre el tema.

En varios lugares y por diferentes medios he tenido la suerte de poder discutir sobre el tema de mi investigación con otras personas con las que también estoy en deuda: los doctores Fernando Rampérez Alcolea, Miguel Ángel Sebastián, Patrick Stokes, Richard Menary y Jonathan Simon; y los no doctores Juan Gómez Bárcena, Kelvin McQueen, Aliosha Celeste Barranco López y Armando Lavalle Terrón.

Tampoco puedo dejar de agradecer el apoyo de la Dra. Ángeles Jiménez Perona a través de mi participación en los proyectos de

investigación «Metaescepticismo y el presente de la epistemología: Postwittgensteinianos y neopopperianos» (HUM2007-60464/FISO) y «Normatividad y praxis: El debate actual después de Wittgenstein» (FFI2010-15975).

Mi gratitud también se extiende a la Universidad Complutense de Madrid, la University of Reading, la Australian National University y la Universidad Nacional Autónoma de México, instituciones a las que tengo que agradecer los medios que pusieron a mi disposición. No me olvido de los investigadores y empleados de las mismas, que me facilitaron enormemente la labor.

Especial mención merece el apoyo económico que recibí de la Universidad Complutense de Madrid a través de una Beca Complutense Predoctoral y de dos ayudas complementarias para realizar estancias en el extranjero.

Por último, quisiera darles las gracias:

A los amigos y compañeros de doctorado que me han soportado hablar (casi) de un único tema durante los últimos cinco años.

A mi novia, quien a pesar de no querer escribir esta tesis por mí ha sido un gran apoyo mientras lo hacía.

A mi familia por sus ánimos y comprensión estos últimos treinta años. Gracias a ellos he podido llegar hasta donde estoy.

INTRODUCCIÓN

«Yo sé quién soy».

Una afirmación que sin duda suscribirían muchos de los filósofos que desfilarán por este trabajo. Quizá sin caer en la cuenta de la carga literaria que la sentencia tiene para casi cualquier lector de lengua hispana.

El caballero andante más famoso de la historia, víctima del primero de sus tropiezos antes incluso de empezar sus aventuras, está siendo acompañado de vuelta hasta su casa por un humilde labrador vecino suyo. Aún sin saber dónde se encuentra, don Quijote confunde al labriego con el marqués de Mantua, uno de los personajes de un romance, y a él mismo con su sobrino Valdovinos. El recién nombrado marqués, aún sin estar del todo convencido de la locura del hidalgo, trata de hacerle entrar en razón:

—Mire vuestra merced, señor, pecador de mí, que yo no soy don Rodrigo de Narváez, ni el marqués de Mantua, sino Pedro Alonso, su vecino; ni vuestra merced es Valdovinos, ni Abindarráez, sino el honrado hidalgo del señor Quijana. (Cervantes, 1605, p. 122)

En ese momento en que la realidad trata de imponerse, podemos imaginar a nuestro héroe, malherido pero orgulloso, levantando la cabeza para hacerle frente:

—Yo sé quién soy —respondió don Quijote—, y sé que puedo ser, no solo los que he dicho, sino todos los doce Pares de Francia, y aun todos los nueve de la Fama, pues a todas las hazañas que ellos todos juntos y cada uno por sí hicieron se aventajarán las mías. (Cervantes, 1605, pp. 122-123)

«Yo sé quién soy» ha sido el *leit motiv* de la discusión sobre la identidad personal en el ámbito de la filosofía analítica desde mediados del

siglo XX. Y a partir de él se han intentado responder algunas de las preguntas más importantes que la filosofía puede plantearse; pues del tema de la identidad penden cuestiones como las que atañen a nuestra supervivencia o a la responsabilidad moral.¹ ¿Cómo saber si seguiré vivo mañana si no sé quién soy? ¿De qué actos deberé responder? ¿Del futuro de quién me debo preocupar? Sin duda sus investigaciones habrían ido mejor encaminadas si desde el principio hubiesen dado más cabida a lo que la literatura podía aportar al tema.

Pero tampoco podemos quejarnos. Ya hace décadas que en la filosofía analítica se abrieron camino las teorías narrativas de la identidad. Y de su mano vinieron las comparaciones entre las narraciones que (dicen) nos constituyen como personas y las que constituyen a los personajes literarios. Comparaciones que en las últimas fechas han sido puestas en entredicho y que aquí pretendo defender junto a un desarrollo personal de lo que denomino el nuevo planteamiento narrativo (NPN).

El actual debate en torno a la identidad personal tiene sus orígenes a mediados del siglo XX cuando autores como Bernard Williams (1957), Sydney Shoemaker (1959), Derek Parfit (1971), John Perry (1972) o David Lewis (1976) recogieron el planteamiento hecho por Locke en su *Ensayo sobre el entendimiento humano*. Para Locke «persona» es un término –que en este contexto debe ser entendido como sinónimo de «Yo»²– que se refiere a una entidad mental distinta del ser humano en cuyo cuerpo está encarnada.³ Su identidad se define en términos de la continuidad de una misma conciencia, siendo la memoria el elemento capaz de unificar ésta a través de las interrupciones cotidianas que sufre a causa del sueño o de pequeños periodos de olvido (Locke, 1690, pp. 318-319). Lo que llama la atención de la postura de Locke es que considera

-
- 1 Para más información sobre a lo que se refiere el problema de la identidad personal véase el anexo I, que debe leerse como una extensa nota al pie de esta introducción.
 - 2 El término «Yo» es equivalente al inglés «*self*». La mayúscula únicamente tiene la intención de permitir al lector diferenciarlo del pronombre personal «yo».
 - 3 Es importante señalar que el debate sobre el problema de la identidad personal en los términos que aquí me van a interesar es totalmente ajeno al dualismo. De hecho aunque Locke admite la existencia del alma, considera que la identidad personal no depende de su continuidad (Locke, 1690, p. 319). En este sentido el término «encarnada», que aparecerá con cierta frecuencia a lo largo de este trabajo, no presupone en ningún caso la existencia previa de una entidad separable del cuerpo que posteriormente se encarna en éste. Las propuestas de solución al problema de la identidad personal a partir del dualismo no serán contempladas.

que cuando nos preocupamos por nuestra supervivencia o por cuestiones de responsabilidad moral –lo que Marya Schechtman denomina las *cuatro características*⁴ de la identidad personal (Schechtman, 1996, pp. 136-162)–, lo que nos interesa es nuestra existencia como personas, y no como seres humanos. En este sentido si la conciencia de un ser humano fuese interrumpida de tal modo que no recordase un crimen que realizó en el pasado, éste no debería serle imputable, pues en rigor habría sido cometido por una persona distinta a la que actualmente habita en el cuerpo del ser humano en cuestión.

La teoría de Locke, a pesar de sus problemas, tiene grandes virtudes. Sin embargo fue entendida por sus críticos como una teoría simplista basada en la continuidad de la memoria, lo que tuvo dos consecuencias. La primera, es que se reveló como patentemente falsa ya desde el siglo XVIII, recibiendo duros ataques que no fueron respondidos en siglos (*vid.* Butler, 1736; Reid, 1785). La segunda, es que el criterio propuesto por Locke para determinar la identidad de una persona basado en la continuidad de la memoria fue entendido por sus lectores en el siglo XX como un criterio *objetivo* que se podía contraponer a otro criterio objetivo: el de la continuidad corporal. Así, ignorando la dimensión subjetiva de la conciencia cuya continuidad era la que definía para Locke la existencia de las personas (*cf.* McDowell, 1997), el debate que se dio desde mediados del siglo XX hasta aproximadamente la década de 1980 estuvo dominado por lo que denomino el planteamiento objetivo (PO).

Todos los participantes del debate que se dio en el PO aceptaron la premisa de Locke de que el término «persona» debía utilizarse para referir a la entidad que nos importa cuando nos preocupamos por las cuatro características. Sin embargo los dos bandos contendientes discutían sobre cuál era dicha entidad. Los partidarios de los criterios físicos –Williams fue su defensor más significativo– defendieron que «persona» era sinónimo de «ser humano», por lo que nuestra identidad debía definirse en términos de la continuidad de nuestro cuerpo como sucede con el resto de los animales. Pero sin duda los partidarios de los criterios psicológicos –entre los que podemos destacar a Shoemaker, Lewis, Parfit y Perry– representaron la posición dominante. Para estos autores, herederos de la postura lockeana, el término «persona» debía ser considerado sinónimo de «Yo» y referirse por tanto a lo que somos en tanto entidades mentales. No

4 Schechtman señala efectivamente cuatro características: la preocupación prudencial por nuestro futuro, la compensación futura por las situaciones presentes, la supervivencia y la responsabilidad moral. Sin embargo creo que las dos primeras están lo suficientemente relacionadas con el problema de la supervivencia como para no considerarlas de forma independiente.

obstante, distanciándose del filósofo inglés, defendieron además que las personas debían ser definidas en términos de su continuidad psicológica (y no la de su conciencia), como consecuencia de la lectura objetivista que hicieron de la teoría de Locke. Su argumentación partía del hecho de que, aunque normalmente la continuidad física y la psicológica coinciden, si imaginamos un caso hipotético en el que divergiesen –por ejemplo debido a un trasplante cerebral– lo que nos importa cuando nos preocupamos por las cuatro características parece ser la continuidad psicológica y no la física.

El debate del PO llegó a su punto álgido (y en cierto sentido a su fin) con la publicación de *Razones y personas* de Parfit en 1984. En su libro el filósofo de Oxford defendió una vertiente de la postura neo-lockeana, mostrando tanto sus virtudes como sus limitaciones. Pues aunque mantiene junto a sus copartidarios que lo que nos debe importar cuando nos preocupamos por nuestra supervivencia es la continuidad psicológica, también argumenta de manera convincente que ésta no se encuentra necesariamente relacionada con la identidad personal. El motivo es que debido principalmente a su vinculación con la responsabilidad moral y nuestra existencia en sociedad, la identidad personal debe ser una relación determinada. Esto es, siempre debe haber una respuesta del tipo sí o no a la pregunta por la identidad de una persona. Dado que la continuidad psicológica es una cuestión de grado que además puede darse en forma ramificada –dos personas futuras podrían tener al mismo tiempo continuidad psicológica con una única persona pasada–, no puede ser una relación equivalente a la de identidad personal, aunque frecuentemente ambas se den de forma conjunta. En consecuencia Parfit argumenta que lo que nos debería importar cuando nos preocupamos por nuestra supervivencia no debería ser que seamos nosotros los que sobrevivamos, sino que sería suficiente para alegrarnos que existiese alguien que se nos pareciese lo suficiente. Es decir, que hubiese alguien psicológicamente continuo con nosotros aunque no tuviese nuestra identidad. En este sentido se entiende el mantra que Parfit venía repitiendo desde tiempo atrás: la identidad personal no es lo que importa (Parfit, 1984, p. 458).

La extravagancia de esta afirmación, y sobre todo su inapelabilidad dentro del marco del PO, hizo que a partir de los años ochenta un nuevo enfoque empezase a abrirse camino en el debate sobre la identidad personal. Se trata de lo que denomino el planteamiento narrativo clásico (PNC), entre cuyos defensores podemos destacar a Alasdair MacIntyre (1981), Jerome Bruner (1987), Daniel Dennett (1988), Charles Taylor (1989), Paul Ricoeur (1990) y Marya Schechtman (1996). Las teorías de estos autores se puede interpretar que reformulan la continuidad psicológica del

planteamiento objetivo en términos de la continuidad de la narración que contamos sobre nosotros mismos, por lo que en cierto sentido heredan los presupuestos de la línea neo-lockeana (*vid.* Slors, 1998). Sin embargo lo hacen a partir de dos críticas al PO que les servirán como base para desarrollar las principales tesis sobre las que se sustenta el PNC. La primera crítica se refiere al hecho de que una de las propiedades más relevantes de nuestra existencia es que poseemos conciencia, un fenómeno radicalmente subjetivo que no se puede reducir a hechos objetivos como por ejemplo la continuidad psicológica. La segunda señala que las cuestiones de responsabilidad moral vinculadas al concepto de identidad personal no tienen sentido fuera del espacio social en el que surgen, por lo que la primacía que los neo-lockeanos daban en el PO a la opinión de la propia persona a la hora de decidir sobre su identidad les incapacitaba para capturar la dimensión social de nuestras vidas. En el fondo ambas críticas tenían un mismo objetivo: señalar que el marco objetivista del planteamiento anterior era incapaz de dar cuenta de ambos aspectos de nuestra existencia por considerarlos *reducibles* a una base física o psicológica objetiva (no en vano Parfit denomina «reduccionista» a su propia postura).

Para solucionar el primero de los problemas y atender a la dimensión subjetiva de la conciencia los teóricos del PNC defendieron lo que siguiendo a Galen Strawson podemos denominar la tesis psicológica de la Narratividad (*vid.* Strawson, 2004). Esta tesis dice que las personas experimentan subjetivamente su vida en forma narrativa, por lo que la continuidad psicológica postulada por los neo-lockeanos en el PO como criterio de identidad debía ser sustituida por la continuidad de la narración con la que comprendemos nuestra vida. Por su parte el segundo problema —el de la imposibilidad de capturar la dimensión social de nuestra existencia si se define a las personas únicamente a partir de lo que éstas piensan sobre sí mismas— trató de solucionarse con la tesis ética de la Narratividad, que señala que comprender la totalidad de la propia vida en forma de una narración es algo necesario para tener una vida buena y desarrollarse plenamente como persona. La idea era que al comprenderse a sí mismas de este modo las personas tendrían en cuenta el contexto en el que se desarrollan sus acciones y por tanto estarían dando cabida a la vertiente social de su existencia. Para asegurarse de que esto fuese así algunos autores como MacIntyre o Ricoeur postularon la necesidad de que las narraciones de la propia persona entrasen en diálogo con las narraciones que los demás hacían sobre ellas, de modo que su identidad fuese el resultado de esta interacción (MacIntyre, 1981, p. 263; Ricoeur, 1990, pp. 161-162). Sin embargo la postura dominante a este respecto fue

la de Schechtman, quien simplemente postuló la necesidad de que nuestra narración sobre nosotros mismos se plegase, al menos hasta cierto punto, a los sucesos reales en los que estuvo involucrada la persona (Schechtman, 1996, pp. 119-130). De este modo para la mayoría de los partidarios del PNC nuestra identidad se constituía únicamente mediante nuestra propia narración sobre nosotros mismos, lo que provocó que la tesis ética de la Narratividad fuese incapaz de solucionar del todo el problema que le dio origen. Pero a pesar de ello, las teorías de la identidad narrativa fueron acogidas con gran entusiasmo, y desde los años ochenta hasta fechas recientes fueron aceptadas casi sin encontrar oposición.

No obstante el entusiasmo inicial, en los últimos años el PNC ha sido objeto de numerosas críticas que finalmente han provocado su abandono. Simplificando al máximo el sentido de estas objeciones, es posible agruparlas en dos bloques. El primero, que podemos considerar encabezado por el artículo de Strawson «Against Narrativity» al que ya me referí en el párrafo anterior (Strawson, 2004), centra su atención en el hecho de que si nuestra identidad está constituida por una única narración, es imposible que ésta atienda al mismo tiempo a todas las facetas de nuestra existencia. El segundo, cuyo principal exponente es Peter Lamarque (2004, 2007), acusa a los narrativistas de haber utilizado el concepto de «narración» de forma ambigua, presuponiendo que toda narración es literaria y abusando así de forma injustificada de las comparaciones entre las narraciones que supuestamente constituyen a las personas y las que constituyen a los personajes literarios.

Este segundo bloque de críticas tiene sin duda gran importancia, y algunos de sus aspectos que aún no han recibido respuesta ocuparán mi atención en el último capítulo. Sin embargo es el primer conjunto de críticas el que ha recibido más atención y el que finalmente provocó el colapso del PNC. Y es que como Schechtman puso de manifiesto al tratar de responder a Strawson, lo que sus críticas revelan es que existe una tensión subyacente en el planteamiento narrativo clásico entre las demandas que hacen sobre la narración que nos constituye las dimensiones subjetiva y social de nuestra existencia (Schechtman, 2007, pp. 168-169). Si intentamos que nuestra narración se adapte lo más posible a nuestra experiencia subjetiva del mundo, acabamos teniendo que renunciar a cualquier sentido significativo de «narración» que nos permita atender a nuestra existencia social. De hecho, algunas personas dicen no experimentar sus vidas en absoluto como narraciones —el propio Strawson asegura ser un ejemplo de ello—, por lo que si el PNC fuese cierto no tendríamos más remedio que, contra toda intuición, denegarles la categoría de «persona». Si por otro lado intentamos que la narración sobre nuestra

vida sea lo suficientemente compleja como para dar cuenta de su dimensión social, nos vemos obligados a renunciar a dar cabida al modo en que subjetivamente perciben sus vidas algunas personas. Siendo así, el PNC se encuentra con la imposibilidad de proponer una única entidad narrativa que posea al mismo tiempo las propiedades necesarias para poder dar cuenta de las dimensiones fenomenológica y social de nuestra existencia.

Ante esta aporía, la solución que ofreció Schechtman constituye el fin del PNC y la base sobre la que desarrollar el nuevo planteamiento narrativo (NPN): acabar con la sinonimia entre «Yo» y «persona». Para Schechtman la mejor alternativa para escapar de los problemas revelados por las críticas de Strawson consiste en suponer que no sólo somos las dos entidades que se habían presupuesto en el debate desde sus inicios –Yoes mentales y seres humanos físicos– y que el término «persona» debía utilizarse para referirse a aquella de las dos que nos debería importar cuando nos preocupamos por las cuatro características. Por el contrario deberíamos considerar que el término «persona» tiene su propio referente, distinto de los de «Yo» y «ser humano», y que este referente es una entidad vinculada de alguna manera no suficientemente especificada con el ámbito social. De este modo el NPN se caracteriza por la defensa de que somos al mismo tiempo tres entidades distintas –seres humanos, Yoes y personas–, al menos una de las cuales –la persona– se constituye de forma narrativa.

Sin embargo, una vez que hemos decidido que el término «persona» ya no se va a utilizar para designar a la entidad que nos importa cuando nos preocupamos por las cuatro características sino para referirse a una entidad cuya existencia no se había contemplado hasta el momento, ya no queda claro cuál de las tres entidades que somos es la que nos debe importar cuando nos preocupamos por nuestra supervivencia o por la responsabilidad moral. De hecho los escasos autores que hasta la fecha han contribuido al NPN –especialmente Schechtman (2007, 2011), Dan Zahavi (2007) y Richard Menary (2008)– tienen opiniones contradictorias al respecto. En cualquier caso no podía suceder de otra manera, pues debido a su reciente surgimiento nadie se ha preocupado todavía de discutir el tema para desarrollar en profundidad este nuevo planteamiento en el que aún ni siquiera hay consenso sobre si los Yoes se constituyen narrativamente o no y sobre cuál es la relación de las personas con el ámbito social.

Como ya anuncié al principio de la introducción, desarrollar el NPN será precisamente el objetivo de estas páginas. Ello supondrá aceptar como punto de partida la solución que Schechtman ofreció como respuesta al primer bloque de críticas recibidas por el PNC. Sin embargo, me distanciaré de ella en dos puntos. El primero es que dado que uno de los

motivos por los que en el NPN se decide diferenciar entre Yo es y personas es la necesidad de dar un lugar apropiado a la dimensión social de nuestra existencia, para elaborar mi concepto de «persona» recurriré, en lugar de a la caracterización que Schechtman dio del concepto, a las sugerencias que MacIntyre y Ricoeur hicieron sobre el papel que las narraciones de los otros tienen en la constitución de nuestra identidad. El segundo, es que defenderé que nuestras intuiciones nos invitan a considerar que las cuatro características de la identidad personal de las que hablaba Schechtman no se aplican como un grupo a una sola de las tres entidades que se distinguen en el NPN. Como trataré de demostrar, cuando nos preocupamos por nuestra supervivencia nos importan en la misma medida los futuros del ser humano, del Yo y de la persona que somos. Sin embargo veremos que cuando hablamos de responsabilidad moral únicamente nos importan las personas, lo cual no es de extrañar teniendo en cuenta que el concepto de «persona» se define en el NPN para atender específicamente a nuestra vida social.

Una vez defendidos estos dos puntos, mostraré que ambos son incompatibles con un marco materialista, por lo que abogaré por la renuncia al mismo. Ésta evidentemente será una de las reivindicaciones más polémicas de mi exposición, pues hasta fechas recientes el materialismo, en sus diversas formas, ha sido prácticamente la única postura aceptada en el ámbito de la filosofía analítica de la mente. No obstante debemos tener en cuenta que la resistencia de las cualidades mentales a ser explicadas en un marco materialista estándar —el famoso *explanatory gap* que parece forzarnos a aceptar una postura eliminativa (Levine, 1983)— ha provocado que en las últimas décadas cobren fuerza algunas posturas alternativas, como el emergentismo de Antonio Damasio (Damasio, 2003, p. 302) o el panpsiquismo de Strawson (1994) y David Chalmers (1996). Independientemente de su valor, lo que estas posturas revelan es la necesidad filosófica de encontrar una alternativa al materialismo estándar. Pero como trataré de demostrar, ninguna de ellas tiene la consistencia suficiente como para considerarse un marco apropiado en el que resolver ni el problema de la conciencia ni tampoco para justificar nuestras intuiciones acerca de que cuando nos preocupamos por nuestra supervivencia nos preocupan en la misma medida las tres entidades que somos. Siendo así, espero que mi apuesta por el abandono de un marco materialista sea vista como una opción no sólo plausible sino también deseable.

En cualquier caso mi intención no será defender algún tipo de dualismo como alternativa al materialismo. Un proyecto así sin duda estaría destinado al fracaso. En su lugar presentaré una propuesta que

retomará una línea de pensamiento que puede considerarse wittgensteiniana: el problema mente-cuerpo es un falso problema derivado de la aplicación errónea de unos conceptos –los referidos a propiedades mentales– en un marco que les es completamente ajeno –el materialista–, por lo que no existe ninguna posible explicación física de la conciencia. El motivo de esta incapacidad sin embargo no será ontológico, como parecen sugerir los emergentistas y los pansiquistas, sino epistemológico.

El principal problema para defender mi postura será la creencia generalizada de que el materialismo no sólo es una explicación satisfactoria de la realidad, sino que es la única posible, ya que su adecuación a la misma no es únicamente a un nivel epistemológico, sino también ontológico. Es decir, la realidad no sólo *se explica* mediante el materialismo, sino que *es* tal y como el materialismo supone, por lo que si los conceptos referidos a las propiedades mentales no son aplicables en un marco materialista entonces no existen en absoluto. Por este motivo en primer lugar argumentaré que ningún discurso científico puede proponerse criterios realistas a la hora de elaborar su marco conceptual, sino sólo criterios pragmáticos, pues su adecuación a la realidad será siempre incontrastable (*vid.* Quine, 1948). Como consecuencia de esto mostraré que según la actitud interpretativa que adoptemos ante esta realidad objetiva e incognoscible –en adelante REALIDAD–, podremos formar diversas explicaciones válidas –esto es, útiles– que sean incompatibles entre sí (*cf.* Dennett, 1991b). De este modo cada una de estas explicaciones presupondrá la existencia de una serie de entidades distintas e irreducibles a un marco común, por lo que podremos decir que cada una de ellas postulará la existencia de un mundo o realidad subjetiva diferente (*vid.* Goodman, 1978). Por supuesto ninguno de estos mundos tendrá una existencia *objetivamente* REAL, pero serán todo lo reales que podemos llegar a conocer. La REALIDAD es incognoscible, por lo que ninguna de las afirmaciones que podamos hacer tendrá nunca un valor ontológico objetivo, sino únicamente relativo a nuestra experiencia subjetiva del mundo.

Una vez defendida mi postura con respecto a los problemas ontológicos y epistemológicos del conocimiento científico mostraré que no sólo podemos resolver así el problema de la conciencia, sino también justificar nuestras intuiciones acerca de qué nos debe importar en nuestra supervivencia y resolver por tanto el problema de la identidad personal. Para hacerlo únicamente tendremos que presuponer que los conceptos de «ser humano», «Yo» y «persona» pertenecen a tres mundos distintos fruto del uso de tres actitudes interpretativas diferentes: Los seres humanos son entidades físicas definidas a partir de lo que podemos denominar una *actitud de diseño*; los Yoes son entidades fenomenológicas definidas desde

una *actitud intencional*; y las personas por su parte son entidades sociales pertenecientes a un mundo descrito desde lo que denominaré una *actitud colectiva*. De este modo podremos ver que, aunque sus condiciones de supervivencia y las características de su identidad son independientes pues emanan del mundo al que pertenecen, podemos identificarnos en el mismo grado con las tres entidades.

Un aspecto que merece la pena destacar es que la definición más original dentro del NPN será la de las personas. En primer lugar porque la actitud interpretativa desde la que las definiré –la actitud colectiva– no procede directamente de la teoría de Dennett como sucede con las actitudes de diseño e intencional, aunque esté basada en ella (*vid.* Dennett, 1991b). Y en segundo, porque al inspirarme para definir las en las posturas de MacIntyre y Ricoeur acerca del papel de las narraciones de los otros, el resultado será que nuestra identidad como personas se constituirá a través de las creencias colectivas que sobre nosotros existan en nuestra sociedad, creencias que se articularán a través de una *negociación narrativa* entre nuestro propio relato de vida y los relatos que los demás cuentan sobre nosotros. Estas consideraciones darán lugar a lo que denominaré la *tesis social de la Narratividad*, una de las principales aportaciones de este trabajo y que podría considerarse una mejora de la tesis ética de la Narratividad.

Volviendo sobre el modo en que el problema de la identidad personal se resuelve en el NPN, una vez que contemos con unas definiciones apropiadas de «ser humano», «Yo» y «persona» mostraré que precisamente la falta de precisión en estos conceptos fue el origen del problema. Así por ejemplo los participantes del debate que se dio en el PO, al no distinguir apropiadamente entre las características de cada entidad, trataron de adscribir una propiedad social como es el tener una identidad determinada ora a los seres humanos ora a los Yoes, cuyas identidades son necesariamente una cuestión de grado. Por su lado la mayor parte de los teóricos del PNC, al igual que los neo-lockeanos del PO, consideraban que nuestra identidad se definía únicamente desde la primera persona –el «yo sé quien soy» del que hablaba al principio de esta introducción–, sin tener en cuenta que aunque el hecho de que el hidalgo Alonso Quijana haya perdido la razón y crea ser un personaje de alguna novela de caballerías puede provocar que ya no sea el Yo que era anteriormente, aún así seguirá siendo el mismo ser humano que era antes de que perdiese la cabeza y ni todos los discursos del mundo podrían convencer a su vecino Pedro Alonso –ni, dicho sea de paso, a nadie en su sano juicio– de que debería concederle los honores que para sí mismo reclama el que ha decidido rebautizarse como don Quijote.

Cuando haya expuesto mi desarrollo del NPN hasta este punto espero que la argumentación presentada sea suficiente para considerar resuelto el problema de la identidad personal. Sin embargo el segundo bloque de críticas que recibió el PNC, referido al supuesto abuso de la comparación entre personas y personajes en base a un ambiguo uso del concepto de «narración», también es aplicable al NPN. Siendo así resultará necesario responder a estas críticas para otorgar una mayor solidez a mi tesis.

La solución partirá del reconocimiento del *problema de la ambigüedad* en el PNC, así como de otros cuatro problemas derivados del mismo: el *problema de la trivialidad*, el *de la explicitud*, el *del escepticismo* y el *de la interpretación*. Cuando tratan de demostrar que nuestra identidad se constituye mediante una narración, los teóricos del PNC oscilan entre un sentido cognitivo de «narración» –que emana de nuestra forma de comprender las acciones humanas–, y otro formalista –vinculado a los rasgos formales que debe tener un texto para ser una narración literaria–, de tal modo que únicamente demuestran que nuestro pensamiento es narrativo en el primer sentido pero se permiten dar un paso injustificado y suponer que también debe serlo en el segundo. Sin embargo mostraré que las críticas realizadas por Lamarque parten de la presuposición de que el sentido relevante de «narración» para los teóricos del PNC es el formalista, cuando en realidad es el cognitivo. Si suponemos que esto es así, entonces tanto el problema de la ambigüedad como los otros cuatro problemas derivados pueden resolverse si a la hora de hacer comparaciones entre las personas y los personajes literarios se tiene en cuenta que no se debe definir a estos últimos en términos formalistas, sino desde un marco más próximo al cognitivo como por ejemplo el utilizado por la estética de la recepción (*vid.* Mayoral, 1987; Warning, 1975). Partiendo entonces de la manera en que esta teoría literaria conceptualiza las narraciones propias de las novelas mostraré que las comparaciones entre Yoes, personas y personajes no sólo son posibles, sino también enormemente productivas.

Teniendo lo anterior en consideración, el presente trabajo se puede dividir en dos partes:

La primera, de naturaleza fundamentalmente expositiva, está dedicada a revisar la evolución histórica del debate sobre la identidad personal en la filosofía analítica hasta la aparición del NPN. Está compuesta por dos capítulos. El primero se centra en el análisis del PO desde su surgimiento hasta su abandono y el segundo hace lo propio con el PNC, mostrando los motivos que hicieron aparecer el NPN.

La segunda, que abarca los cuatro capítulos restantes, es más argumentativa y tiene como objetivo responder las críticas recibidas por el PNC y desarrollar el NPN para resolver el problema de la identidad

personal. El tercer capítulo se centra en la defensa de que cuando nos preocupamos por la responsabilidad moral sólo nos importan las personas, que cuando nos preocupamos por nuestra supervivencia nos importan en la misma medida las tres entidades que somos, y que esto último es injustificable en un marco materialista. El cuarto ofrece una alternativa anti-realista al materialismo estándar que nos permitiría solucionar el problema de la conciencia y justificar el hecho de que nos preocupemos en el mismo grado por la supervivencia de seres humanos, Yoes y personas. En el quinto se definen estas tres entidades teniendo en cuenta el marco anti-realista presentado en capítulo anterior, tras lo cual se muestra cómo el problema de la identidad personal –entendido como la pregunta sobre qué y quién somos, en qué consiste nuestra persistencia a lo largo del tiempo y qué entidad o entidades es la que nos debe importar cuando nos preocupamos por las cuatro características– se resuelve sin complicaciones en el NPN. Finalmente en el sexto capítulo se exponen las acusaciones que se lanzaron contra el PNC en lo que a su uso del concepto de «narración» se refiere, respondiéndolas de tal modo que se haga evidente que las comparaciones entre Yoes, personas y personajes en el NPN son tanto posibles como provechosas.

Como decía nada más comenzar, la sentencia «yo sé quién soy» parece haber sido el *leit motiv* de la discusión en torno al problema de la identidad personal en la filosofía analítica desde sus orígenes. Esta presunción ha dado lugar a argumentos basados en lo que cada uno de los participantes del debate tenía la impresión de ser, como he señalado en esta introducción y tendremos ocasión de ver más adelante. Este trabajo pretende escapar de esa tendencia ofreciendo una concepción más amplia de qué es aquello a lo que nos referimos cuando decimos «yo», que incluye lo que cada uno de nosotros creemos ser pero que no se reduce a ello. Espero que la argumentación aquí presentada convenza al lector y le abra nuevas perspectivas en lo que concierne al problema de la identidad personal.

CAPÍTULO 1

EL PLANTEAMIENTO OBJETIVO

Como adelanté en la introducción, a lo largo de las primeras décadas de la segunda mitad del siglo XX el debate sobre la identidad personal en la filosofía analítica se centró en los criterios objetivos que podían utilizarse para determinar qué entidad es la que nos importa cuando nos preocupamos por la supervivencia o por la responsabilidad moral. Partiendo del supuesto de que el término «persona» debía ser utilizado para referirse a dicha entidad, a un lado se encontraban los defensores de que «persona» y «ser humano» debían ser entendidos como sinónimos que se referían a lo que somos en tanto animales, constituidos únicamente por nuestro cuerpo. Al otro, los que siguiendo una tradición lockeana defendían que «persona» –que se consideraba sinónimo de «Yo»– se refería a una entidad mental relacionada con la memoria y la autoconciencia que debía ser distinguida de lo que somos en tanto seres humanos, pues los seres humanos no son un «Yo», sino que *tienen* un «Yo». Como es de suponer, los *animalistas* buscaban un criterio físico para determinar la identidad de las personas, mientras que los *neo-lockeanos* hacían lo propio buscando un criterio psicológico. Sin embargo ninguna de las dos posiciones consiguió su propósito, y las soluciones que unos aportaban no tardaban en ser puestas en evidencia por los otros.

A partir de los años ochenta –especialmente tras la publicación de *Razones y Personas* de Derek Parfit en 1984– la sensación de que este planteamiento objetivo (PO) era erróneo fue en aumento, provocando el desplazamiento del problema de la identidad personal hacia las concepciones narrativas. Y es que como Parfit puso de manifiesto en su

libro, hay ideas que deben mantenerse juntas o caer juntas, y si sostenemos una postura reduccionista¹ los problemas que los experimentos mentales de fusión y duplicación de personas plantean a ambos tipos de criterios nos fuerzan a elegir una de dos opciones: (1) o la identidad personal es una relación indeterminada o (2) depende de una decisión arbitraria o basada en hechos triviales. Sea cual sea la opción que escojamos, la conclusión que Parfit extrae es que la identidad personal no es lo que realmente nos importa cuando nos preocupamos por nuestra supervivencia (*vid.* Parfit, 1984, pp. 371-603). Las tesis de Parfit fueron consideradas inaceptables pese a la inapelabilidad de su argumentación, ya que nadie en el marco del PO parecía dispuesto a renunciar ni al reduccionismo, ni a la creencia de que nuestra identidad es determinada, ni por supuesto a la evidencia de que en nuestra supervivencia es precisamente nuestra identidad lo que está en juego. Siendo así, la búsqueda de un nuevo marco era inevitable.

A pesar de que los enfoques narrativos posteriores –en gran medida herederos de la postura neo-lockeana²– dejan de lado la conclusión parfitiana como parte de un debate abandonado, es interesante observar un poco más de cerca la discusión que se dio en el PO para comprender el sentido del posterior giro narrativo, aclarar algunas de nuestras intuiciones sobre las personas y ayudarnos a crear un concepto de las mismas que nos permita resolver el problema de la identidad personal. Por otro lado, considero que Parfit estaba esencialmente en lo cierto. El reduccionismo nos obliga a aceptar una de las dos conclusiones señaladas, por lo que en

- 1 Se puede considerar que el reduccionismo es el marco ontológico compartido por todos los partidarios del PO. Parfit lo define a partir de la idea de que alguien es reduccionista si considera que una descripción completa del mundo no tiene que incluir más que los hechos físicos y los hechos psicológicos que estén asociados a éstos. Frente a esta postura, a juicio de Parfit los no-reduccionistas tendrían dos opciones: (1) considerar que la identidad personal depende de la existencia de una entidad puramente mental que es completamente separable del cuerpo o (2) considerar que aunque no somos una entidad separable del cuerpo, existe un hecho adicional a la simple continuidad física y psicológica que es el que determina nuestra identidad. Sin embargo también señala que esta segunda opción tiene que ir unida a la primera, pues no puede haber un hecho adicional que no implique la consideración de que somos entidades completamente separables de nuestro cuerpo. En este sentido su presentación del reduccionismo parece ser poco más que una inocente negación del dualismo cartesiano, como ya señalasen Ricoeur (1990, p. 129), McDowell (1997, p. 244) e incluso Shoemaker (1985, p. 135). En cualquier caso para tener una idea más completa de lo que supone el reduccionismo se puede consultar el libro de Parfit (1984).
- 2 En este sentido puede verse como Slors (1998) y Schechtman (2001) conciben la narratividad como un tipo de continuidad psicológica.

última instancia es incompatible con una solución satisfactoria del problema de la identidad personal. Siendo así, espero que mi argumento a favor de abandonar el materialismo se vea reforzado tras revisar este debate histórico.

En cualquier caso, hay algunas cuestiones previas que merece la pena abordar antes de comenzar con el análisis. En primer lugar, aunque podría parecer que la división entre los partidarios de los criterios físico y psicológico responde a la distinción más primitiva entre dualistas y materialistas, es necesario recordar que se trata de un debate estrictamente materialista. De hecho el razonamiento dualista tiene más en común con el criterio físico que con el psicológico, pues defienden que la supervivencia de las personas se debe a la existencia continua de un alma inmaterial, del mismo modo que los animalistas argumentan a favor de la existencia continua de un mismo cuerpo. Como veremos, el criterio psicológico –aunque en algunas de sus versiones comparte con el dualismo la creencia de que podemos sobrevivir a la muerte de nuestro cuerpo actual– tiene que tomar una forma más elaborada ya desde el siglo XVIII.

Otro aspecto a destacar antes de iniciar nuestro repaso es que frente a la claridad de la intuición animalista –somos nuestro cuerpo, un objeto físico– los partidarios del criterio psicológico, ya desde la formulación de Locke, recurrieron a distintos experimentos mentales para defender sus tesis. Situaciones fantásticas o de ciencia-ficción que trataban de poner en cuestión nuestras intuiciones sobre aquello que somos. Los partidarios del criterio físico respondieron a su vez con nuevos experimentos que reforzasen su intuición primaria, de modo que más que un debate de ideas elaboradas se puede considerar que fue una discusión entre intuiciones que posteriormente se defendieron con argumentos contruidos al efecto –recordemos la tesis de Kripke de que no puede haber mayor argumento a favor de una posición que su contenido intuitivo (Kripke, 1980, p. 50)–. Por este motivo organizaré este capítulo no en función de qué autor defendió tal o cual postura, sino a partir de los principales experimentos mentales que aparecen en las páginas de sus trabajos, a menudo situados a medio camino entre la filosofía y la ciencia-ficción.

Un último punto que merece nuestra atención es precisamente la polémica que el método de los experimentos mentales en este campo despertó entre algunos autores (*vid.* Wilkes, 1988). Y es que aunque se trata de situaciones lógicamente posibles cuya irrealizabilidad es meramente técnica –podemos imaginar un futuro en el que sí sean posibles– se ha argumentado que intentar aplicar un concepto como el de «persona» a una situación completamente nueva no puede producir resultados fiables, ya que un concepto no tiene fuerza lógica más allá de

las situaciones para las que fue creado (Quine, 1972, p. 490). Es decir, no sabemos realmente cómo debería aplicarse el término en contextos ajenos al de su uso normal. Este argumento de Quine, muy relacionado con su tesis de la traducción radical, se basa en la relatividad ontológica inherente a todo lenguaje, pues pese a las apariencias nuestros conceptos nunca están definidos más allá de nuestras necesidades comunicativas. Alguien podría responder que existe «algo» en el mundo real que son las personas, y es a ese «algo» al que nos referimos al hablar, por lo que la investigación por medio de los experimentos mentales es un método perfectamente válido. Es una forma de descubrir cuál es el referente real del término «persona». Sin embargo, tal y como Quine mostró, la idea de «referencia» es un sinsentido a no ser que se entienda como relativa a un sistema de coordenadas previo, a un lenguaje de fondo. Si en ese lenguaje de fondo el objeto que estamos tratando de identificar no está definido no hay ningún modo de determinar la referencia (*vid.* Quine, 1968). Como consecuencia, lo que somos en tanto personas no está determinado más allá de nuestra experiencia cotidiana. Al mismo tiempo, y justo por esta indeterminación de los conceptos, el «resultado» de los experimentos –lo que nos sentimos inclinados a creer tras leerlos– está altamente condicionado por las descripciones que se hagan de los mismos. Por poner un ejemplo, pensemos en dos descripciones de un mismo caso de teletransportación. En la primera utilizaremos la tercera persona:

Una persona entra en un cubículo de teletransporte en la Tierra. Aprieta el botón y el escáner comienza a funcionar. En primer lugar la deja inconsciente, y a continuación recorre de arriba a abajo su cuerpo pulverizándolo y registrando a la vez la posición de todas las moléculas que lo componen. Al terminar no queda nada en el cubículo aparte de un montoncito de polvo y un fuerte olor a carne quemada. Entonces la información recogida por el escáner es enviada por radio a un punto de teletransporte situado en Marte. Allí, utilizando un depósito de moléculas orgánicas, una máquina construye un nuevo cuerpo que será idéntico al de la persona original, aunque estará constituido por partículas distintas. Cuando termine de construirlo, ese cuerpo que antes era materia inerte cobrará vida y afirmará ser la misma persona que minutos antes entró en la máquina de teletransporte en la Tierra.

En la segunda, reproduciré una parte de la descripción que hace Parfit del mismo caso:

Aunque creo que esto es lo que va a suceder, todavía vacilo. Pero entonces recuerdo cómo se reía mi mujer cuando, hoy al desayuno, le manifesté mi nerviosismo. Como me recordó, ella ha sido teletransportada a menudo, y nada va mal con *ella*. Aprieto el botón. Como se me pronosticó, pierdo la conciencia

y aparentemente la recobro enseguida, pero en un cubículo diferente. Examinando mi cuerpo, no encuentro ningún cambio en absoluto. Hasta está todavía en su sitio el corte que me hice en el labio superior esta mañana al afeitarme. (Parfit, 1984, pp. 371-372)

Como el lector habrá notado, a pesar de no afirmarlo explícitamente, la primera descripción sugiere que el escáner en realidad está matando a la persona que entra en su interior, reduciéndola a un montón de polvo. Que posteriormente se construya una réplica no cambia el hecho de que la persona original ha muerto. Más que una forma de viajar, el teletransporte parece una forma de matar y producir clones en otro lugar del mundo. Por el contrario la segunda descripción, al focalizar el relato desde la conciencia del usuario, crea la sensación de que el teletransporte realmente está trasladando a la persona de un lugar a otro. Sin embargo podríamos decir coherentemente que en realidad hay dos sujetos en la historia que se confunden al estar escrita en primera persona. Uno es responsable de apretar el botón y acto seguido muere. El otro cobra vida en Marte creyendo sin fundamento que unos minutos antes se encontraba en la Tierra cuando en realidad no existía con anterioridad. En consecuencia tenemos que admitir que el caso del teletransporte –al igual que todos los demás experimentos mentales– no *demuestra* nada en absoluto. Descrito de formas distintas incita al lector a llegar a conclusiones opuestas porque no sabemos aplicar el concepto de «persona» en dicho caso. No podemos saber qué pasa con las personas cuando se teletransportan porque no sabemos qué podría significar que una persona sobreviviese (o muriese) en una situación así.

Sin embargo, aunque es cierto que los experimentos mentales no demuestran nada, al leerlos no sentimos que dichas situaciones sean indescriptibles o ajenas. Más bien al contrario, provocan en nosotros fuertes reacciones. Movilizan nuestras creencias y sacan a la luz lo que intuitivamente pensamos que pasaría en casos que de hecho podrían llegar a suceder. Por eso considero que aunque los experimentos mentales no pueden utilizarse para determinar cuál pudiera ser el *verdadero* significado de «persona», sí que se pueden utilizar para sacar a la luz nuestras intuiciones y ayudarnos a crear una definición que sea lo más coherente posible con ellas, pues como señalé en la introducción el problema precisamente es que el concepto carece de un significado definido. Una vez aclarado este punto, podemos comenzar con el análisis.³

3 Para ampliar información se puede recurrir al libro *Personal Identity* editado por John Perry (1975), que además de tener una interesante introducción al tema recoge una buena selección de los textos que discutiré en este apartado.

1.1. El príncipe y el zapatero

Se puede considerar como el experimento que da inicio a la oposición entre los criterios físico y psicológico de la identidad personal una breve reflexión que presenta Locke en el capítulo 27 del libro II de su *Ensayo sobre el entendimiento humano*. En ella nos dice que si un príncipe muriese y la conciencia de su vida pasada se introdujese en el cuerpo de un zapatero, a pesar de que príncipe y zapatero son dos seres humanos distintos, ambos serían la misma persona (Locke, 1690, pp. 323-324). El experimento es fácil de entender, pues encontramos múltiples ejemplos de este tipo de historia en la literatura universal. Es la idea de que el alma de una persona puede introducirse en otro cuerpo y no obstante seguir siendo la misma. Sin embargo la originalidad de Locke reside en desvincular el suceso de la idea de alma para centrarse en la conciencia, estableciendo la memoria como el elemento capaz de unificar ésta a través de las interrupciones provocadas por el sueño y el olvido. Esto es así porque mientras la identidad de los seres humanos está determinada por la posesión de un cuerpo y un alma que transmite vida a las distintas partículas de materia que pasan a formar parte de ese cuerpo, las personas son una entidad mental distinta de ambos. Aunque nuestros pensamientos no fuesen tenidos siempre por una misma sustancia –es decir, aunque tuviésemos distintas almas a lo largo de la vida–, éstas quedarían unidas en una única persona por participar de una misma conciencia, del mismo modo que las distintas partículas que forman parte del cuerpo de un animal a lo largo de su historia quedan unidas por participar de una misma vida continuada. Por su parte, una misma alma podría animar a distintos seres humanos en distintas épocas, pero si no llevase la conciencia de la vida de una a la de la otra, se trataría de personas diferentes. Al ser la memoria el elemento unificador de la conciencia, recordar mis actos pasados es lo que los hace míos y demuestra que soy la misma persona que los realizó. En este sentido podemos afirmar que no es lícito castigar a una persona por un delito que no recuerde haber cometido, pues en rigor no fue *ella* la que lo hizo.

Como vimos en la introducción, la teoría de Locke fue interpretada por sus críticos como una teoría simplista basada en la continuidad de la memoria en vez de en la continuidad de la conciencia, lo que tuvo dos consecuencias importantes. La primera, es que ya en el siglo XVIII se puso de manifiesto que la teoría presentaba graves problemas que fueron objeto de duras críticas que no se pudieron rebatir en siglos. La segunda,

Para revisiones más críticas pueden consultarse también los libros de Marya Schechtman (1996) y Mariano Rodríguez González (2003).

es que el criterio de continuidad de la memoria fue entendido por los lectores de Locke en el siglo XX como un criterio *objetivo* que se podía contraponer a otro criterio objetivo, el de la continuidad física. De este modo la dimensión subjetiva, consciente, de nuestra existencia, que era la base de la teoría de Locke, fue ignorada en el PO (*vid.* McDowell, 1997). Pero será mejor no adelantarnos a nuestros pasos. Como decía, las críticas a la postura de Locke no se hicieron esperar, y ya en el siglo XVIII encontramos dos objeciones de peso que no encontrarían respuesta en casi doscientos años. Una es de Joseph Butler (1736) y señala lo que se conoce como el problema de la *circularidad*. La otra es de Thomas Reid (1785) y se refiere al problema de la *transitividad*.

La idea de Butler es que la memoria no puede ser el criterio de la identidad personal porque por definición sólo podemos recordar nuestras propias experiencias, de modo que la memoria presupone la identidad personal (Butler, 1736, p. 100). El concepto de «memoria» depende lógicamente del de «persona», por lo que no puede ser su fundamento so pena de caer en la circularidad. El argumento es de una sencillez y evidencia tales que rebatirlo fue complicado, e incluso hoy en día algunos pueden considerar extravagante la mejor respuesta que ofrecieron los defensores del criterio psicológico y que fue elaborada por Sydney Shoemaker en base al concepto de «cuasi-memoria» (Shoemaker, 1970, p. 271).

Cuando recordamos algo es debido a que fuimos testigos o agentes del suceso. No nos es posible recordar una experiencia sin situarnos en ella de una u otra manera, porque todos nuestros recuerdos proceden de experiencias que tuvimos nosotros mismos. En eso es precisamente en lo que se basa Locke para considerar la memoria como el elemento unificador de la conciencia. Sin embargo aunque en nuestro caso suceda así, no es lógicamente necesario que lo haga. Podemos imaginar un mundo en el que fuese posible recordar sucesos de los que nosotros no fuimos testigos. La memoria presupone que el agente recordador y el sujeto original de la experiencia recordada coinciden. Pero podemos utilizar el concepto de «cuasi-memoria» para referirnos a recuerdos independientemente de quién fuese el sujeto que tuvo la experiencia original. En este sentido la memoria es simplemente un caso concreto de cuasi-memoria en el que ambos sujetos son el mismo. Esta descripción parece vaga y en un primer momento da la impresión de no referirse a nada en concreto, pero en realidad se trata de una idea que se ha desarrollado con cierta frecuencia en la literatura y el cine de ciencia-ficción, y que podemos entender mejor a la luz de los replicantes de la película *Blade Runner* (Ridley Scott, dir., 1982).

Un replicante es un ser humano –hay ciertas diferencias biológicas, pero irrelevantes en este momento– que ha sido fabricado ya en su forma adulta para realizar un determinado trabajo. Sin embargo un adulto con la mentalidad de un recién nacido no sería útil, por lo que su cerebro se fabrica directamente con una serie de conocimientos incorporados que según para qué se les vaya a utilizar son más o menos amplios. En el caso de los replicantes destinados a ser soldados puede ser suficiente con unos conocimientos básicos de tipo bélico y una personalidad sumisa. Por el contrario si se espera que el replicante interactúe con otras personas sin que éstas sepan diferenciarlos de un ser humano normal, los conocimientos tienen que ser mucho más completos. Es imprescindible que el replicante *crea* ser una persona normal, para lo que es necesario dotarla de los recuerdos de una vida pasada que no tuvo. Como crear una nueva historia vital para cada replicante sería una tarea ingente, el fabricante utiliza los recuerdos de otras personas –en el caso de la historia central de la película, la sobrina del dueño de la empresa– para implantárselos a los nuevos seres –su secretaria, Rachael–. De este modo la replicante Rachael se recuerda a sí misma teniendo unas experiencias que en realidad fueron tenidas por otra persona. Es decir, no las recuerda, sino que las cuasi-recuerda. Cuando Rachael descubre la verdad se encuentra entonces con la imposibilidad de diferenciar entre los cuasi-recuerdos que experimentó otra persona antes de que ella existiese y los cuasi-recuerdos que experimentó ella misma posteriormente, sufriendo una comprensible crisis de identidad. Sus cuasi-recuerdos provienen de dos fuentes distintas y no sabe cuál de las dos es realmente ella. Sin embargo en nuestro caso sabemos que nuestros cuasi-recuerdos proceden de una única fuente, lo que unido al hecho de que el concepto de «cuasi-memoria» es anterior al de «persona» nos da un soporte adecuado para establecer un criterio de identidad que evite el problema de la circularidad señalado por Butler.⁴

Una objeción que se le suele plantear a la idea de «cuasi-memoria» es que para que algo sea realmente un recuerdo tiene que estar relacionado con la experiencia original del modo apropiado. Si tenemos un recuerdo

4 Señala acertadamente Manuel Cruz que en realidad no hace falta echar mano de la ciencia-ficción para ejemplificar el concepto de «cuasi-memoria», pues todos somos susceptibles de asumir como propios recuerdos ajenos, ya sea mediante hipnosis o por la vía más común de confiar en los relatos que otros nos cuentan sobre nuestra infancia: «No parece una situación inverosímil: a todos nos costaría, si nos pusiéramos a ello, distinguir fielmente qué nos fue contado y qué recordamos efectivamente de buena parte de episodios de nuestra primera infancia» (Cruz, 2005, p. 42).

que no se relaciona con una experiencia real –si no fue impreso en nuestra mente a causa de dicho suceso–, no se trata de un recuerdo sino de una alucinación. De este modo se podría decir que los replicantes en realidad no cuasi-recuerdan nada, sino que tienen alucinaciones de un cierto tipo. Sin embargo esta crítica está basada en una concepción de la memoria bastante alejada de la que hoy en día se considera acertada, y es más cercana a la metáfora de la tablilla de cera de Platón que al modelo de red de procesamiento de información en paralelo que propone actualmente la neurociencia.⁵ La metáfora platónica de la memoria nos hace pensar en nuestra mente como en un mecanismo de grabación que registra todos los detalles que la realidad *imprime* en nuestros sentidos. Se trata de un proceso causal típico, lo que permite pensar a los críticos de la idea de «cuasi-memoria» en cuáles son las causas apropiadas para que algo cuente como un recuerdo. Sin embargo en la actualidad nuestra concepción de la memoria se parece más a un proceso de tomar apuntes que nos permitan reconstruir la experiencia cuando nos sea necesario. Pero dado que sólo memorizamos unas líneas generales, cada vez que recordamos un suceso la reconstrucción que elaboramos se solapa con la experiencia primera, de modo que los apuntes «originales» se ven modificados y en las sucesivas ocasiones sólo podemos recordar en base a la última reconstrucción que hicimos (*cf.* Kandel, 2007, p. 327; McCrone, 2003, pp. 26-29). Pondré un ejemplo que pese a su extremismo puede servir como modelo de cualquier recuerdo. Hace un tiempo paseaba con un amigo y surgió una discusión a propósito de *Pórtico*, una novela de ciencia-ficción escrita por Frederik Pohl que ambos habíamos leído unos años atrás. El final de la historia es una consecuencia lógica de algunos supuestos básicos de la teoría de la relatividad, que en su momento nos pareció muy bien utilizada. Sin embargo, a pesar de que los dos recordábamos el episodio de forma parecida, un análisis superficial mostró que nuestro recuerdo era incoherente con lo que sabíamos de la teoría de la relatividad, lo que nos hizo suponer que debíamos estar recordando mal el final de la obra. Tras varias horas de debate –a lo que contribuyó tanto nuestra inclinación a la vida ociosa como el hecho de que nuestro conocimiento de la teoría de Einstein es sólo superficial– pudimos reconstruir el argumento de la novela de forma que fuese coherente con lo que sabíamos. Ahora ambos recordamos el final de la obra de una forma que creemos que es más fiel al texto, sin embargo no se puede decir que lo hagamos debido a nuestra

5 Se puede encontrar un interesante análisis de las diferentes metáforas que se han propuesto para explicar el funcionamiento de la memoria a lo largo de la historia y de sus implicaciones en el libro *Las metáforas de la memoria* (Draaisma, 1995).

experiencia original de lectura. De hecho ya no recordamos el final que teníamos en la memoria antes de nuestra discusión y no sabemos si nuestra reconstrucción es acertada o si hemos elaborado un final alternativo. La conexión de nuestro recuerdo de la novela con la lectura original es más lejana después de recordarla, pues ha aumentado la incertidumbre sobre aquello que hemos distorsionado. En este sentido analizar la memoria en términos de causalidad con la experiencia original se vuelve algo falaz. Recordar se parece más a inventarse el pasado que a observarlo de nuevo, haciendo que la diferencia entre la memoria y el tipo de alucinación que podría ser un cuasi-recuerdo se difumine.⁶

Volviendo a las críticas al argumento de Locke, la segunda de las objeciones de peso que recibió su versión del criterio psicológico fue la realizada por Reid, quien señaló que a diferencia de la identidad, la memoria no es una propiedad transitiva:

Supongamos a un valiente oficial que fue azotado en la escuela cuando era un niño por robar en un huerto, que robó un estandarte al enemigo en su primera campaña, y que fue ascendido a general en su madurez: Supongamos también, lo que hay que reconocer que es posible, que cuando consiguió el estandarte era consciente de haber sido azotado en la escuela, y que cuando ascendió a general era consciente de su toma del estandarte, pero había perdido completamente la conciencia de su azotaina.

Si suponemos estas cosas, se sigue de la doctrina del señor Locke que la que fue azotada en la escuela es la misma persona que tomó el estandarte, y que la que tomó el estandarte es la misma persona que fue nombrada general. De donde se sigue, si hay algo de verdad en la lógica, que el general es la misma persona que aquella que fue azotada en la escuela. Pero la conciencia del general no llega tan lejos como para recordar su azotaina, por lo tanto, según la doctrina del señor Locke, no es la persona que fue azotada. Por lo tanto el general es, y al mismo tiempo no es, la misma persona que fue azotada en una escuela. (Reid, 1785, pp. 114-115)⁷

6 Esta objeción al concepto de cuasi-memoria se relaciona con tres versiones del criterio psicológico que se diferencian en lo que consideran el tipo correcto de causa para poder hablar de continuidad psicológica –causa normal, fiable o cualquier causa–. Volveré sobre esto en el apartado «El señor Brownson».

7 La traducción es mía. El original es:

«Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.

En vista de este argumento, la única forma de que el criterio de Locke fuese correcto pasaría por un mundo en el que fuese imposible que una persona olvidara una sola de las experiencias que tuvo a lo largo de su vida. Como esto es manifiestamente falso, debemos rechazar el criterio psicológico de Locke, pues al ignorar la no-transitividad de la memoria lleva a la contradicción de afirmar que una persona al mismo tiempo es y no es la misma.

A pesar de la importancia de esta objeción, fue respondida con solvencia y no supone ya ningún problema. La solución tiene dos partes. La primera consiste en modificar el criterio de la memoria para convertirla en una relación transitiva y sortear el problema de los olvidos propios de una vida normal. La segunda consiste en incluir otros tipos de conexiones psicológicas, como el carácter, los deseos, las creencias, etc. (Parfit, 1984, pp. 381-383). De este modo, aunque se produjese una amnesia completa, si se conservasen éste otro tipo de conexiones podríamos decir que se trata de la misma persona. Esta segunda parte es fácilmente comprensible y no necesita mayor explicación. Sin embargo la primera requiere un poco más de atención por nuestra parte, pues la misma lógica debe aplicarse al resto de conexiones psicológicas para que eviten también el problema de la transitividad.

El criterio de identidad propuesto por Locke, tal y como fue interpretado por sus críticos y los partidarios del PO, se basaba en la existencia de conexiones directas y concretas de memoria, de modo que dos personas que existiesen en momentos diferentes serían la misma si una fuese capaz de recordar directamente lo que experimentó la otra. Llamaremos a esta relación «*conexividad*». Pero nuestra vida mental no está compuesta de unos pocos recuerdos, sino de incontables. Cada día se forman cientos o miles de recuerdos, de los cuales sólo unas decenas pervivirán más allá de unos cuantos días, por lo que la *conexividad* es una relación que admite grados. Es máxima cuando recordamos todo lo que hicimos en un determinado momento y mínima cuando no recordamos nada. De este modo, aunque no recordemos qué comimos ayer, existen multitud de recuerdos directos que nos permiten identificarnos con la

These things being supposed, it follows, from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging, therefore, according to Mr Locke's doctrine, he is not the person who was flogged. Therefore the general is, and at the same time is not the same person with him who was flogged at a school.»

persona que éramos entonces, por lo que diremos que hay una conexividad fuerte. Si vamos hacia atrás en el tiempo, recordamos cada vez menos cosas, por lo que el grado de conexividad es menor. Sin embargo aunque no tengamos muchos recuerdos directos de lo que hicimos el martes pasado, el miércoles sí los teníamos. Y del mismo modo, aunque no recordemos absolutamente nada de lo que hicimos hace veinte años, cada día entre entonces y la actualidad hemos tenido multitud de recuerdos directos de lo que sucedió el día anterior. Diremos entonces que existe una *continuidad* de memoria con la persona que éramos hace veinte años basada no en la existencia de recuerdos directos, sino en la de cadenas parcialmente superpuestas de conexividad fuerte. Si aplicamos el mismo razonamiento al resto de conexiones psicológicas como las creencias, los deseos y el carácter, podemos formular el concepto de «*continuidad psicológica*» que, a diferencia de la conexividad lockeana, sí es una relación transitiva y nos permite elaborar un criterio psicológico de identidad que evite el problema señalado por Reid.

1.2. Los dobles de Guy Fawkes

Una crítica realizada ya en el siglo XX a la validez de los criterios basados en la memoria es que cualquier entidad psicológica depende en última instancia de una referencia al cuerpo físico para poder ser individualizada, paso previo imprescindible para poder hacer juicios de identidad. Este argumento fue esgrimido por Bernard Williams en 1957 basándose en una versión revisada del experimento del príncipe y el zapatero cuya repercusión en el debate posterior sobre la identidad personal sería enorme al poner de relevancia el problema de la *duplicación*.⁸ Según Williams, podemos imaginar a un hombre que de repente se despierta un día y sus recuerdos han sido sustituidos por aquellos que poseía Guy Fawkes⁹ antes de morir. No sólo sus recuerdos,

8 El problema de la duplicación ya había sido señalado en el siglo XVIII por Reid (1785, p. 114), pero fue el argumento de Williams el que lo puso en el primer plano del debate. Por otro lado Schechtman entre otros, al centrarse en el criterio psicológico, considera este tercer problema una derivación del viejo problema de la transitividad (Schechtman, 1996, p. 30). Sin embargo, aunque tienen la misma estructura lógica ($A=B$, $A=C$, pero $B \neq C$), como veremos más adelante el problema de la duplicación plantea también serias dudas sobre la viabilidad de los criterios físicos, por lo que merece ser considerado de forma independiente.

9 Guy Fawkes un un personaje histórico bastante conocido en Inglaterra. Fue arrestado el 5 de noviembre de 1605 por planear un atentado contra el

sino que también su carácter, su personalidad y todas sus características mentales son idénticas a las de este traidor inglés del siglo XVII. Siguiendo el criterio psicológico, dicho hombre dejaría de ser la persona que había sido hasta entonces y pasaría a ser Guy Fawkes. Hasta aquí todo es prácticamente igual al caso de Locke y no presenta mayores problemas. Sin embargo al mismo tiempo en la casa contigua otro hombre se despierta víctima del mismo suceso paranormal: él también posee todos los recuerdos de Guy Fawkes. Si no hay ningún problema lógico que nos impida imaginar el primer caso, tampoco lo hay en el segundo, por lo que siguiendo de nuevo el criterio psicológico nos vemos forzados a admitir que tenemos a otro Guy Fawkes. Pero como la identidad es necesariamente una relación uno-uno,¹⁰ no es posible que ambas personas sean Guy Fawkes aunque sean cualitativamente idénticas (Williams, 1957, p. 239).

Por definición, los dos hombres tienen los mismos argumentos para reclamar la identidad personal de Guy Fawkes, pero sólo uno de ellos –como máximo– puede estar en lo cierto. Al no tener base para tomar una decisión, según nuestra versión del criterio psicológico la identidad de estos dobles es indeterminada. Esta incapacidad para ofrecer una respuesta no puede eliminarse aportando nueva información, ya que conocemos todos los datos relevantes. La pregunta en este caso sencillamente no tiene respuesta. Podríamos afirmar que cada uno de ellos tiene la mitad de la identidad de Guy Fawkes –signifique lo que signifique esto–, pero convertir la identidad personal en una cuestión de grado no evita que se transforme en una relación indeterminada. Una respuesta a la situación que sí nos permitiría conservar la creencia de que la identidad personal es determinada sería ofrecer una respuesta basándonos en hechos triviales –recordemos que, como señalo en el anexo I, la indeterminación de la identidad es relativa a un marco conceptual dado–, por ejemplo que el hombre que estuviese más cerca de la reina de Inglaterra es el auténtico Guy Fawkes y el otro es sólo una persona muy parecida. Sin embargo una respuesta de este tipo sería arbitraria y no significaría nada, pues sería incapaz de justificar la importancia que atribuimos a la identidad personal (Williams, 1970, pp. 178-179). En consecuencia la única respuesta posible

parlamento británico. Sin embargo aunque confesó su participación en el plan, se negó a dar los nombres de sus cómplices. Fue ejecutado por traición, y desde entonces cada 5 de noviembre se celebra en el Reino Unido la noche de Guy Fawkes, como conmemoración del fracaso del atentado. Su nombre ha adquirido una mayor popularidad internacional en los últimos años debido a que el protagonista de la película *V de Vendetta* (Andy y Larry Wachowski, dirs., 2006) utiliza una máscara inspirada en la que los niños ingleses utilizaban para celebrar la noche de Guy Fawkes.

10 Véase el anexo I «Sobre la relación de identidad».

en este caso, si queremos conservar la determinación de la identidad, es que ninguno de los dos hombres implicados es Guy Fawkes. Pero al mismo tiempo, y por extensión, si sólo existiese un hombre que se hubiese despertado con los recuerdos de Guy Fawkes tampoco podría reclamar su identidad, pues ésta no puede depender de un hecho también trivial y externo a la relación entre ambos como el que no haya nadie más en su misma situación.¹¹

El criterio psicológico lleva a confusión en este caso al no poder diferenciar entre identidad cualitativa –los dobles de Guy Fawkes son cualitativamente idénticos al original en lo que se refiere a sus propiedades psicológicas– e identidad numérica, lo que demuestra su falsedad. Es más, no sabemos qué podría ser identificar numéricamente una entidad psicológica si no es mediante referencia al cuerpo físico que la experimenta, por lo que si eliminamos el cuerpo de la ecuación eliminamos con él todo contenido de la idea de identidad personal (Williams, 1957, p. 241). Aunque se propusiera una nueva versión corregida, el caso de la duplicación de Guy Fawkes muestra que cualquier criterio psicológico dependerá en última instancia de uno físico para individualizar a la persona.

1.3. El señor Brownson

El problema de la duplicación puso en jaque a los defensores del criterio psicológico. Sin embargo sus efectos no sólo fueron devastadores para los neo-lockeanos, sino que a la larga también se mostró como un argumento definitivo en contra de las concepciones animalistas, iniciando el camino que finalmente llevaría al giro hacia la identidad narrativa (Rodríguez González, 2003, pp. 92-93). El primer paso en este sentido fue

11 Este principio sobre la relación intrínseca que un ente debe mantener consigo mismo para poder hablar de identidad es generalmente aceptado, aunque ha habido defensores del criterio psicológico, como Shoemaker, Parfit o Nozick, que han renunciado a él. De este modo llegan a afirmar que la identidad personal depende de la continuidad psicológica únicamente si no hay más de un candidato a ser la misma persona. No obstante considero que esta posición es débil y no la contemplo como una respuesta válida al problema de la duplicación (*cf.* Schechtman, 1996, pp. 30-32). Por otro lado, parece difícil defender por qué una situación de supervivencia exitosa, como es el hecho de que un hombre se despierte con los recuerdos de Guy Fawkes, debería ser considerada como un fracaso cuando se da por duplicado, en lugar de como un doble éxito. Volveré sobre esto más adelante.

la crítica a los criterios físicos elaborada por Shoemaker en torno a la posibilidad de un trasplante cerebral:

Así pues, imaginemos lo siguiente. En primer lugar, supongamos que la ciencia médica ha desarrollado una técnica mediante la cual un cirujano puede extraer completamente el cerebro de una persona de su cabeza, examinarlo u operar en él, y luego ponerlo de nuevo en el cráneo (reconectando los nervios, vasos sanguíneos, etc.) sin causar la muerte ni lesiones permanentes; vamos a imaginar que esta técnica de «extracción cerebral» ha llegado a ser una práctica generalizada en el tratamiento de tumores cerebrales y otros trastornos del cerebro. Un día, para comenzar nuestra historia, un cirujano descubre que un asistente ha cometido un terrible error. Dos hombres, el Sr. Brown y el Sr. Robinson, han sido operados de sendos tumores cerebrales, y en ambos casos se utilizó la técnica de extracción cerebral. Al final de las operaciones, sin embargo, el asistente sin querer puso el cerebro de Brown en la cabeza de Robinson, y el cerebro de Robinson en la cabeza de Brown. Uno de los pacientes muere inmediatamente, pero el otro, el que posee el cuerpo de Robinson y el cerebro de Brown, finalmente recupera la conciencia. Llamemos a este hombre «Brownson». Al recuperar la conciencia Brownson muestra una gran conmoción y sorpresa ante la visión de su cuerpo. Entonces, al ver el cuerpo de Brown, exclama incrédulo «¡Ese que está ahí tumbado soy yo!». Señalándose a sí mismo dice «¡Este no es mi cuerpo; es aquel de allí!». Cuando se le pregunta su nombre, automáticamente responde «Brown». Reconoce a la esposa y a la familia de Brown (a quienes Robinson nunca había conocido), y es capaz de describir detalladamente acontecimientos de la vida de Brown, haciéndolo siempre como si fuesen acontecimientos de su propia vida. De la vida pasada de Robinson no muestra ningún conocimiento en absoluto. Durante un período de tiempo se le observa y se concluye que posee todos los rasgos de la personalidad, los gestos, los intereses, los gustos, y así sucesivamente, que habían caracterizado a Brown, y que actúa y habla de una forma totalmente ajena a como lo hacía el antiguo Robinson. (Shoemaker, 1963, pp. 23-24)¹²

12 La traducción es mía. El original es:

«So, let us imagine the following. First, suppose that medical science has developed a technique whereby a surgeon can completely remove a person's brain from his head, examine or operate on it, and then put it back in his skull (regrafting the nerves, bloodvessels, and so forth) without causing death or permanent injury; we are to imagine that this technique of "brain extraction" has come to be widely practiced in the treatment of brain tumors and others disorders of the brain. One day, to begin our story, a surgeon discovers that an assistant has made a horrible mistake. Two men, Mr. Brown and Mr. Robinson, had been operated on for brain tumors, and brain extraction had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men

El caso es de nuevo una variante de la historia del príncipe y el zapatero. Y como entonces, la gran mayoría de nosotros se inclinaría a pensar que el superviviente de la operación es Brown. Pero a diferencia de los experimentos anteriores, en éste tenemos un elemento físico subyacente —el cerebro— lo que nos lleva a pensar en la posibilidad de una respuesta afirmativa también por parte de los defensores del criterio físico. En principio la situación podría describirse como un trasplante de cerebro, en el que Robinson es el receptor de un órgano que antes era de Brown. En la actualidad se realizan trasplantes médicos de prácticamente cualquier parte del cuerpo humano sin que se vea afectada la identidad del receptor. Se pueden trasplantar riñones, pulmones, hígados, corazones, huesos, piernas, brazos... Y en ninguno de estos casos se considera la posibilidad de que el donante esté «trasladándose» de un cuerpo a otro, sino que se dice que el órgano trasplantado pasa a formar parte del cuerpo del receptor. Sin embargo esto es así porque ninguno de dichos órganos, aunque en muchos casos sean vitales, tiene la importancia que tiene el cerebro. Los órganos trasplantados pasan a estar bajo las órdenes de un nuevo cerebro, que es el que parece determinar nuestra identidad. Podríamos pensar que se trata de una cuestión de número —el receptor sigue siendo la misma persona porque la mayoría de órganos que posee el cuerpo resultante siguen siendo los que eran suyos originalmente—, pero aún en el caso de que el número de órganos trasplantados fuese mayor que el de los órganos originales del paciente, seguiríamos diciendo que el receptor conserva su identidad, pues dichos órganos pasan a estar controlados por el cerebro de éste. De este modo creo que es mejor describir el caso propuesto por Shoemaker no como un trasplante de cerebro, sino como una situación límite en la que el señor Robinson es el donante y el señor Brown ha recibido un trasplante de cuerpo entero. Si aceptamos que esto es así, entonces los defensores del criterio físico no

immediately dies, but the other, the one with Robinson's body and Brown's brain, eventually regain consciousness. Let us call the latter "Brownson". Upon regaining consciousness Brownson exhibits great shock and surprise at the appearance of his body. Then, upon seeing Brown's body, he exclaims incredulously "That's me lying there!" Pointing to himself he says "This isn't my body; the one over there is!" When asked his name he automatically replies "Brown". He recognizes Brown's wife and family (whom Robinson had never met), and is able to describe in detail events in Brown's life, always describing them as events in his own life. Of Robinson's past life he evidences no knowledge at all. Over a period of time he is observed to display all of the personality traits, mannerisms, interests, likes and dislikes, and so on that had previously characterized Brown, and to act and talk in ways completely alien to the old Robinson.»

pueden apelar a la continuidad de todo el cuerpo para hablar de la supervivencia de las personas, sino como máximo a la existencia continua del cerebro.¹³

Este papel determinante que el criterio físico tiene que otorgar al cerebro lo acerca a su vez al criterio psicológico y ofrece apoyo a la intuición de que lo importante es la continuidad psicológica, pues sin duda si le damos tanta importancia a este órgano es por ser la base física de nuestra vida mental. Si pudiésemos separar la continuidad del cerebro y la continuidad psicológica, entonces deberíamos primar ésta última. En este sentido podemos hablar de dos extremos dentro de las variantes del criterio psicológico en función del papel que le asignan a la existencia física continuada del cerebro en su relación con la continuidad psicológica: la versión restringida y la amplia.¹⁴ La versión restringida defiende que para poder hablar de continuidad psicológica es estrictamente necesario que ésta sea producida por una causa normal, lo que implica no sólo la existencia continua del mismo cerebro, sino también que nuestros recuerdos y demás propiedades psicológicas estén conectadas del modo apropiado con experiencias pasadas reales. Si la continuidad psicológica es consecuencia por ejemplo de un recuerdo falso –no producido por la experiencia a la que parece referirse, sino inducido por una situación posterior–, no es una continuidad real, sino aparente.¹⁵ De este modo la versión restringida del criterio psicológico coincide en líneas generales con la versión modificada del criterio físico que acabamos de ver, y ambos aceptarían que Brownson es Brown porque el cerebro determina la identidad, pero que el zapatero de Locke no es la misma persona que era el príncipe, dado que aunque existe una aparente continuidad psicológica, ésta no se ha producido por una causa normal.

Por su parte la variedad amplia del criterio psicológico defiende que podemos hablar de continuidad sea cual sea la causa que la produzca, aún a riesgo de que en algunos casos esto pueda suponer la pérdida de la identidad personal. Parfit lo explica del siguiente modo. Supongamos que unos científicos desarrollan unos ojos artificiales que mediante un microordenador son capaces de conectarse a los nervios ópticos de una

13 A este respecto Parfit discute la posición de Thomas Nagel, quien defiende que efectivamente lo que somos como personas es aquello que causa nuestra continuidad psicológica. Esto es, nuestro cerebro (*vid.* Parfit, 1984, pp. 486-487).

14 Como digo, se trata de extremos, por lo que entre medias existen distintas variantes. En cualquier caso éste hecho no afecta a la exposición general que estamos realizando.

15 Véase la discusión a propósito de los cuasi-recuerdos en el apartado «El príncipe y el zapatero».

persona ciega y devolverle la visión. Las experiencias visuales que producen estos ojos son cualitativamente idénticas a las de los ojos normales, pero sin embargo la causa de dichas experiencias es un fenómeno extraño que implica el procesamiento de las ondas lumínicas por parte del microordenador. Alguien podría afirmar entonces que al no producirse por la causa normal, las personas que utilizan ojos artificiales en realidad no tienen visión, sino una pseudovisión. Sin embargo, aunque estrictamente hablando no serían capaces de ver, lo que tendrían sería *tan bueno como* ver. De igual manera puede que alguien cuya continuidad psicológica con una persona pasada dependa de una operación neurológica no tenga la identidad de dicha persona, pero lo que tiene es *tan bueno como* la identidad personal (Parfit, 1984, pp. 383-386). Esto lleva a Parfit a defender su famosa tesis de que la identidad personal no es lo que importa en nuestra supervivencia. La idea es que si la identidad personal nos parece relevante sólo es porque normalmente va unida a la continuidad psicológica. De este modo tenemos un hecho que no importa –esto es, la identidad personal– pero que normalmente va acompañado de un hecho psicológico que sí importa, por lo que adquiere una importancia prestada. Considerando entonces que la importancia que le atribuimos a la identidad personal procede de la continuidad psicológica, si se produce ésta pero no se da la primera debemos admitir que se ha conservado lo que realmente importa, aunque ello suponga renunciar a la identidad personal. Pero volveré sobre esto un poco más adelante. Ahora sólo me interesa señalar que para el criterio psicológico amplio Brownson es Brown y el zapatero de Locke es el príncipe. O como mínimo, que el zapatero tiene todo lo que debería haber preocupado al príncipe acerca de su supervivencia.

Pero como decía anteriormente, lo que complica la discusión sobre la identidad personal son los casos de duplicación. Ya hemos visto que el criterio psicológico es vulnerable a ellos, por lo que si el criterio físico resultase airoso deberíamos aceptar que cualquier discusión desde el punto de vista psicológico es estéril, incluidas las reflexiones sobre el hecho de que la continuidad del cerebro sólo es importante por ser la base de nuestra vida mental y la más famosa de Parfit sobre la no importancia de la identidad. El criterio psicológico restringido evitaría el problema de la duplicación y sería válido únicamente por la referencia a un criterio físico –la continuidad del cerebro– al que en última instancia sería reducible; y el criterio psicológico amplio sería simplemente erróneo, pues no podría ofrecer una respuesta satisfactoria. Sin embargo, el experimento del señor Brownson abre una grieta dentro de la invulnerabilidad del criterio físico al problema de la duplicación.

Una vez aceptado que el criterio físico sólo puede apelar a la continuidad del cerebro, el problema al que se tiene que enfrentar es al hecho de que no es necesaria la existencia de *todo* el cerebro para mantener a un ser humano con vida. En casos de epilepsia se llega incluso a extraer por completo uno de los hemisferios cerebrales del paciente, lo que demuestra que la existencia de medio cerebro es suficiente para la supervivencia de la persona. Pero si un ser humano puede sobrevivir con la mitad del cerebro y seguir siendo la misma persona, en principio sería lógicamente posible utilizar la otra mitad para implantarla en el cuerpo de otro hombre (*vid.* Parfit, 1984, pp. 454-467). Para alejarnos de los casos de epilepsia en los que la mitad cerebral extraída está gravemente dañada, podemos imaginar una variante del experimento del señor Brownson. Tres hermanos gemelos, Charles, John y George, sufren un accidente de tráfico. El cuerpo de Charles queda gravemente dañado, pero su cerebro está intacto. Por su parte, los cerebros de John y George tienen lesiones mortales irreparables, pero sus cuerpos están en buen estado. Como la técnica del trasplante cerebral tiene una probabilidad de éxito del veinte por ciento, los médicos deciden cortar el cerebro de Charles por la mitad e implantar la parte derecha en el cuerpo de John y la izquierda en el de George para aumentar sus probabilidades de supervivencia. Sin embargo, en contra de las estadísticas, ambos trasplantes funcionan y Charles recobra la conciencia en dos nuevos cuerpos. Como según el criterio físico revisado medio cerebro es suficiente para reclamar la identidad, en ese caso tanto Charles-John como Charles-George tendrían aspiraciones legítimas a ser Charles, lo que es incompatible con las propiedades básicas de la relación de identidad. Consecuentemente, igual que en el caso de los dobles de Guy Fawkes, la única respuesta posible a esta situación es defender que ninguno de los dos sería George, pues la duplicación excluye la identidad (Williams, 1960, p. 47). Pero por extensión una persona que perdiese la mitad del cerebro tampoco podría continuar siendo ella misma aunque la otra mitad no estuviese viva en otro cuerpo, lo que parece ir en contra de nuestras intuiciones más básicas.

Una operación del tipo descrito en este experimento está sujeta a ciertas objeciones que deben ser tenidas en cuenta, pues podrían invalidar las conclusiones. La primera y más importante es la duda acerca de si dividir un cerebro en dos mitades provocaría la existencia de dos flujos de conciencia independientes. Si no fuese así, no podríamos asegurar que los hombres que sobreviven a la operación fuesen personas, pues la conciencia ha sido tradicionalmente una de las características más importantes para ser considerado como tal (*vid.* Dennett, 1976, p. 9; Strawson, 1959, p. 104). Sin embargo se trata de un aspecto que ha podido

ser comprobado –todo lo que se puede comprobar acerca de la existencia de experiencias conscientes– mediante el análisis del comportamiento de pacientes a los que por motivos médicos se les había seccionado el cuerpo calloso, que es la parte del cerebro que mantiene unidos ambos hemisferios.

Un experimento real tipo que muestra el desdoblamiento de la conciencia en estos casos consiste en presentar al sujeto una pantalla ancha, en la que el lado izquierdo es azul y el lado derecho es verde, de forma que cada ojo sólo puede ver uno de los dos colores. En cada mitad de la pantalla se muestra sobreimpresa una pregunta. «¿Cuántos colores ves?». Se le pide al sujeto que escriba con ambas manos la respuesta a la pregunta que ve escrita, y en los dos casos la respuesta es la misma: «Sólo uno». Acto seguido se cambia la pregunta. «¿Qué color ves?». Para mayor sorpresa, una de las manos contesta «verde» y la otra «azul». Al responder de este modo, no hay motivos para dudar de que el sujeto está teniendo sensaciones visuales conscientes tanto del color verde como del color azul. Sin embargo cada una de las experiencias procede de un único ojo y por tanto se dirige a un único hemisferio cerebral. Al tener seccionado el cuerpo calloso no existe comunicación entre las dos partes del cerebro, que se manejan entonces como si fuesen dos esferas distintas de conciencia (Parfit, 1984, pp. 441-444). De este modo no cabe duda de que si el experimento de los tres hermanos gemelos es posible, los dos supervivientes tendrían experiencias conscientes diferentes que estarían producidas por un único hemisferio cerebral.

La segunda objeción al experimento de los hermanos se refiere a la diferencia en el funcionamiento de ambos hemisferios, pues lo normal en el caso de los adultos es que se dediquen a funciones distintas. Así por ejemplo el habla suele estar controlada por el hemisferio izquierdo, mientras que las capacidades musicales suelen estar más desarrolladas en el derecho. Sin embargo este tipo de problema, aunque podría afectar a algunas de las versiones del criterio psicológico –podría interrumpirse la continuidad– no afecta en el caso del criterio físico. Al mismo tiempo, las versiones más desarrolladas del criterio psicológico prestan atención a varios tipos de relaciones psicológicas, por lo que aunque al dividir el cerebro se interrumpiese la conexividad de un cierto tipo de propiedades, podrían seguir existiendo otras que fuesen suficientes para garantizar la continuidad.

Finalmente, una última objeción se refiere a la imposibilidad de dividir el cerebro inferior sin dañarlo de modo que no pueda llevar a cabo ninguna de sus funciones, lo que podría provocar que el caso fuese impracticable. Al tratarse de una parte del sistema nervioso imprescindible para el

mantenimiento de funciones vitales como la respiración o el movimiento cardíaco, si al seccionarlo ninguna de las dos mitades funcionase el trasplante no podría mantener con vida al paciente. Sin embargo se trata de una mera imposibilidad técnica que podría llegar a ser evitada en un futuro. Por ejemplo podría desarrollarse un órgano artificial que supliese las funciones básicas que hubieran sido afectadas por la división, de modo que esta objeción tampoco afecta a la validez del experimento.

El caso de los tres gemelos pone en evidencia que el problema de la duplicación afecta tanto al criterio físico como al psicológico, llevándonos a deducciones que consideramos inaceptables. Una solución en ambos casos sería concluir que la duplicación elimina la identidad personal, pero a costa de aceptar que la relación de identidad de una persona consigo misma depende de un elemento extrínseco a la relación entre ambas como es la no existencia de una tercera persona. Al mismo tiempo parece difícil justificar por qué una situación que debería ser considerada como supervivencia –el trasplante de la mitad de mi cerebro a otro cuerpo–, si se diese por duplicado debería considerarse como una defunción en lugar de como una doble supervivencia –las dos mitades de mi cerebro son trasplantadas–. Otra posibilidad sería modificar nuestros criterios de identidad para poder ofrecer una respuesta a los casos de duplicación, pero dado que las dos personas resultantes tienen exactamente la misma relación con la persona original, cualquier decisión se basaría en un hecho trivial incapaz de justificar la importancia que le damos a la identidad. Finalmente, una tercera respuesta sería dar cabida a la posibilidad de que la identidad sea una relación indeterminada y que en los casos de duplicación sencillamente no exista una respuesta a la pregunta sobre la supervivencia. Sin embargo la creencia de que nuestra identidad es determinada es una de las intuiciones más profundas que subyacen al debate –véase el anexo I–, por lo que renunciar a ella es una opción que por el momento no estoy dispuesto a considerar.

En los dos siguientes apartados argumentaré todavía más a favor de las ideas de que tanto el criterio físico como el psicológico deben o bien hacer depender la identidad de un hecho trivial o bien aceptar que es indeterminada. Pero antes de ello merece la pena señalar, aunque sea brevemente, una respuesta alternativa al problema de la duplicación que podría adaptarse para apoyar a cualquiera de los dos criterios, aunque no ha tenido una aceptación suficiente.

La solución consiste en considerar a las personas como objetos que sólo existen a través del tiempo, es decir, como entidades cuatridimensionales. En este sentido en nuestra vida diaria cuando decimos ver a una persona en realidad estamos viendo una fase-de-persona (en inglés «*person-stage*»

o «*time-slice*»), que es sólo una parte temporal de la persona en cuestión. Las personas son objetos de cuatro dimensiones formados por una serie de fases-de-persona unidas de la forma apropiada. Aunque el argumento se puede utilizar para defender del problema de la duplicación tanto al criterio físico como al psicológico, sus mayores defensores han sido partidarios de este último (Lewis, 1976; Perry, 1972) por lo que para estos autores la forma apropiada en la que deben estar unidas las fases-de-persona es la continuidad psicológica. En cualquier caso, si modificamos este punto y decimos que la forma apropiada en la que deben estar relacionadas varias fases-de-persona para constituir una persona es la continuidad del cuerpo, el argumento es perfectamente aplicable al criterio físico.

Para resolver el problema de la duplicación desde esta postura sólo tenemos que observar que una fase-de-persona no tiene por qué ser parte de una única persona, sino que podría darse el caso de una fase-de-persona compartida. Podemos imaginar a dos siameses unidos por un dedo, de forma que entre los dos únicamente tienen tres pulgares. Si no se los separa, el pulgar compartido es a la vez parte de dos personas distintas. De forma análoga una fase-de-persona podría ser compartida por dos personas que no tienen cuerpos separados en ese momento temporal, pero que lo tendrán —o lo han tenido— en otro momento (Perry, 1972, pp. 468-469). Así por ejemplo, en el caso de los tres gemelos, antes del accidente las fases-de-persona de Charles no pertenecían a una única persona, sino que eran compartidas por Charles-John y Charles-George. En contra de lo que pueda parecer, Charles no era una persona sino un conjunto de fases-de-persona compartidas por dos personas que aún no se habían separado, por lo que el accidente era necesario para mostrar que Charles había sido siempre parte de dos personas distintas. De este modo sólo observando la vida al completo de una persona podemos saber si sus fases-de-persona han sido todas exclusivamente suyas o si por el contrario ha habido momentos en los que han sido compartidas con alguien más.

En cualquier caso, a pesar de solucionar el problema de la duplicación, la extravagancia de afirmar que cualquiera de nosotros podría ser en realidad dos personas que aún no se han separado ha hecho que esta postura no haya tenido el peso suficiente como para salvar ninguno de los criterios de identidad personal que se manejaron en la segunda mitad del siglo XX en el marco del PO, por lo que considero que ya se ha dicho suficiente.¹⁶

16 Se puede encontrar una exposición y defensa del cuatridimensionalismo en Bordes Solanas (1998).

1.4. El científico malvado

Como hemos visto, el experimento del señor Brownson sembró dudas sobre la validez del criterio físico. Al mostrar que sólo puede apelar a la continuidad del cerebro nos hace pensar que lo que de verdad importa en la supervivencia de las personas es la continuidad psicológica. De este modo la defensa de que la continuidad física del cerebro es necesaria parece sostenerse únicamente en el hecho de que es la base de nuestra vida mental, por lo que si fuese posible separarlas la continuidad del cerebro ya no haría falta. Sin embargo Williams todavía presentó otro experimento mental para defender el criterio físico basándose en nuestros sentimientos hacia un hipotético futuro de nuestro cuerpo (Williams, 1970, pp. 167-168).

Digamos que un científico malvado –yo mismo– decide que dentro de cinco minutos va a secuestrar al lector de estas páginas –sólo al que continúe leyendo después del siguiente punto y seguido– y que mañana le va a someter a una espantosa tortura. El dolor que vas a experimentar, lector, será insufrible, pero todo lo haré por el bien de la investigación sobre el problema de la identidad personal. No tengo especial interés en que sufras de un modo innecesario, por lo que para tranquilizarte te puedo asegurar que no tienes de qué tener miedo. Para cuando comience la tortura no recordarás nada de mi advertencia. Es más, no recordarás absolutamente nada de tu vida anterior. Entiendo que aún sientas reparos. Pero, ¿qué me dices si antes de que empiece la tortura elimino además todos tus deseos, creencias y rasgos de tu personalidad? El dolor, no te quiero mentir, va a ser tan grande que ni siquiera te lo puedes imaginar. Tal vez pienses que si te desmayas dejarás de sentirlo, pero tengo mucha paciencia. Todavía me quedan años antes de dar por concluida mi investigación y puedo esperar a que te despiertes. Incluso puedo esperar a que estés lo suficientemente recuperado de tus heridas como para volver a empezar. Sin embargo no va a existir ninguna continuidad psicológica entre la persona que eres ahora y la que serás mañana. Es más, mientras presiono tus nervios directamente con el objetivo de lograr producirte el máximo dolor posible, no sólo no sabrás que eres tú, sino que habré modificado tus conexiones neuronales para que sean como las de Napoleón. ¿No te alivia saber que creerás ser Napoleón mientras continúo con mi investigación? Sin duda debes de sentirte mejor sabiendo que tendrás el carácter de un hombre tan valeroso mientras desempolvo mis instrumentos medievales. Además, antes de comenzar grabaré toda la información contenida en tu cerebro en el cerebro de otra persona. Digamos en el de ese compañero de tu oficina bajito y feo que tanto odias. A partir de mañana, mientras yo me encargo de que cada minuto del resto

de tu vida sea un infierno de dolor, el hombre bajito y feo creará ser tú. Y seguramente engañe a todo el mundo. Especialmente a tu mujer, pues ya sabes que ella es una firme defensora del criterio psicológico. Como ves, no tienes ningún motivo para preocuparte, pues mientras tu estés sufriendo los dolores más inhumanos que yo sea capaz de infligirte, el hombre bajito y feo que creará ser tú se estará acostando con tu mujer.

A la luz de esta situación hipotética, Williams argumenta que nuestro sentimiento de terror hacia una tortura en el futuro no disminuye por el hecho de que vayan a producirse en nosotros determinados cambios psicológicos. Tú seguirás siendo esa persona porque tu cuerpo seguirá a disposición del científico malvado. Que además te vayas a volver loco y a creer que eres Napoleón, y que otra persona se vaya a volver loca y a creer que es tú, no cambia en absoluto la situación. O si la cambia, es para hacerla todavía más temible. Tú no serás el hombre bajito y feo porque la continuidad psicológica por si misma no afecta a la identidad personal. Por supuesto podría acusármese, igual que se podía acusar a Williams, de manipular las creencias del lector debido a la descripción que he hecho del caso. La conclusión ya está establecida desde el uso que hago de los pronombres, presuponiendo que a pesar de modificar tus recuerdos, tú seguirías siendo torturado. Pero como dije al principio, todos los experimentos mentales utilizan este recurso, y lo que realmente está en juego en la discusión son las intuiciones que los apoyan. Y la intuición a favor de que tenemos motivos para temer la tortura es muy fuerte.

Tras el análisis de este experimento en *Razones y Personas* Parfit decide modificarlo ligeramente para averiguar cómo funcionaría nuestra intuición en casos en los que una persona «se transforma» en otra (Parfit, 1984, pp. 419-440), señalando lo que podemos llamar el problema de la *fusión*. Si aceptamos las conclusiones del experimento de Williams, esto no sucede cuando únicamente se modifican las conexiones psicológicas de una persona A. Si por el contrario sustituyésemos su cuerpo por el de otra persona B, pero organizásemos su cerebro de forma que tuviese continuidad psicológica con A, los neo-lockeanos tampoco aceptarían que la persona original ha desaparecido. Sin embargo si cambiásemos la totalidad del cuerpo de A por el de otra persona B –sin modificar sus conexiones psicológicas–, independientemente del criterio que sigamos tendríamos que aceptar que ya no se trataría de la persona A original. De este modo podemos imaginar que el científico malvado decide que antes de torturarte va a cambiar tu cuerpo por el de Napoleón. Te situará en una máquina de su invención que, en un abrir y cerrar de ojos, eliminará todas tus células y las sustituirá por réplicas exactas a las de Napoleón. Si hiciese esto de golpe, no tendrías ningún motivo para temer la tortura.

Aunque el proceso implique irremediamente tu muerte –cosa que parece más que probable–, tú no serás el que vaya a sufrir los dolores que le seguirán, sino que será el antiguo emperador francés.¹⁷ Sin embargo, ¿qué pasaría si la sustitución no se hiciese de golpe? Supongamos que el primer día introducirá tu cuerpo en la máquina, que sustituirá únicamente el uno por ciento de las células de tu cuerpo y de tu cerebro. El segundo día volverá a repetir el proceso, sustituyendo otro uno por ciento, y así continuará hasta que el centésimo día tu cuerpo haya sido reemplazado en su totalidad por el de Napoleón. Sin duda los primeros días la persona resultante apenas será diferente de ti. Quizá empieces a tener algún recuerdo borroso de la Batalla de Waterloo, y tus manos sean más ásperas, pero es difícil argumentar que un cambio del uno o el dos por ciento pueda producir tu muerte. Por el contrario los últimos días la persona que entre en la máquina será prácticamente idéntica a Napoleón, por lo que nos sentimos inclinados a pensar que después de sustituir el noventa y cinco por ciento de tu cuerpo efectivamente te has transformado en Napoleón. Sin embargo en los casos intermedios no tenemos claro cuál debería ser la respuesta. No parece evidente que haya algún porcentaje crítico de modo que si han sustituido el cincuenta por ciento de tu cuerpo aún sobrevives, pero si sustituyen una célula más te conviertes en Napoleón. La supervivencia de una persona no puede depender de un hecho trivial como es el que una célula de tu dedo meñique ya no sea tuya. Que la célula sea una neurona no cambia nada, pues la supervivencia de una neurona –o de las escasas conexiones psicológicas dependientes de ella– tampoco puede afectar de forma significativa al resultado. Una solución de este tipo sería arbitraria.

El argumento se parece sospechosamente al problema Sorites, que dice que si quitamos un grano de un montón de arena, no podemos transformar el montón en algo que no sea un montón. De este modo, si alguien toma un montón de arena y va disminuyéndolo grano a grano, finalmente, debido a nuestro razonamiento, le quedará un montón de arena formado por un único grano, lo cual es evidentemente absurdo. Sin embargo nuestra conclusión cuando se aplica a las personas no difiere de la respuesta de sentido común al problema Sorites. Cuando han sustituido el noventa y nueve por ciento de tu cuerpo todos estamos de acuerdo en que la persona resultante no eres tú, del mismo modo que un único grano de arena no es un montón. El problema reside en la consideración de los casos intermedios. Alguien podría decir que para que algo sea un montón

17 Si en realidad la persona que sufrirá la tortura será Napoleón o sólo una réplica suya es algo que está sujeto a debate, evidentemente, pero por agilizar el análisis permítaseme suponer que efectivamente es Napoleón.

de arena tiene que tener por lo menos diez granos y a pesar de la arbitrariedad de la decisión nadie se sentiría turbado, pero ya hemos rechazado este tipo de respuesta para el problema de la identidad personal. Del mismo modo alguien puede decir que el concepto de montón es indeterminado, por lo que no se puede establecer cuándo algo deja de ser un montón. En los casos intermedios, la pregunta acerca de si algo es un montón o no es una pregunta vacía, pues nuestro concepto de montón no tiene una significación definida. Pero como hemos insistido a lo largo de este capítulo, la identidad personal tiene que ser una relación determinada. No es posible decir que no hay respuesta a la pregunta sobre nuestra supervivencia, o que sólo sobrevivimos a medias porque nuestra otra mitad pertenece a Napoleón. En consecuencia, ni el criterio físico ni el psicológico pueden ser ciertos a no ser que aceptemos que la identidad es indeterminada o depende de una decisión arbitraria.

1.5. El teletransporte

El último experimento del que voy a hablar es el ya citado del teletransporte, presentado por Parfit en *Razones y Personas* para mostrar de nuevo la fuerza de nuestras intuiciones a favor del criterio psicológico. Al principio del capítulo ya apunté en qué consistía la idea, pero la repetiré brevemente. Una persona, por ejemplo Douglas Quaid, utiliza una máquina de teletransporte para viajar a Marte. Como está previsto, su cuerpo en la Tierra es destruido y posteriormente se construye un cuerpo idéntico en la estación de destino. Cuando Quaid –que podemos imaginar que es un partidario del criterio psicológico amplio–, sale de la máquina de teletransporte en Marte, no le cabe ninguna duda de que es la misma persona que minutos antes se encontraba en la Tierra. Físicamente es incapaz de encontrar ninguna diferencia, y mentalmente la continuidad es absoluta, por lo que la conclusión lógica a la que llega es que el teletransporte es una forma de viajar. Sin embargo, dadas estas premisas, podemos imaginar una continuación de la historia en la que Douglas Quaid se dispone a viajar de nuevo:

Estoy otra vez en el cubículo, listo para otro viaje a Marte. Pero esta vez, cuando aprieto el botón verde, no pierdo la conciencia. Se escucha un zumbido y luego el silencio. Salgo del cubículo y le digo al asistente: «No funciona. ¿Qué hice mal?».

«Sí que funciona», contesta, y me da una tarjeta impresa. Leo: «El nuevo escáner graba un cianotipo de usted mismo sin destruir su cerebro ni su cuerpo. Esperamos que sepa apreciar las oportunidades que este avance técnico ofrece».

El asistente me cuenta que soy una de las primeras personas que usan el nuevo escáner. Añade que si me quedo una hora podré usar el comunicador para verme y hablar conmigo en Marte.

«Un momento», contesto, «si estoy aquí no puedo estar *también* en Marte».

Alguien tose con mucha cortesía, un hombre de bata blanca que me pide hablar en privado conmigo. Nos vamos a su despacho, me dice que me siente y hace una pausa. Luego dice: «Me temo que tenemos problemas con el nuevo escáner. Graba su cianotipo con la misma perfección y exactitud, ya lo podrá comprobar cuando se vea y hable consigo mismo en Marte. Pero parece que resulta nocivo para el sistema cardíaco cuando lo explora. A juzgar por los resultados que hemos tenido hasta ahora, aunque estará usted en Marte con una salud perfecta, aquí en la Tierra tiene que esperar un ataque cardíaco en los próximos días». (Parfit, 1984, p. 372)

A diferencia del caso del teletransporte simple, en esta continuación de la historia aparece el problema de la duplicación. Al estar focalizada además desde la perspectiva de la persona que se queda en la Tierra, la historia refuerza la impresión de que el teletransporte no es una forma de viajar, sino de fabricar réplicas. No sólo eso, sino que también es una forma de matar a la persona original. En el caso del teletransporte simple, la muerte es instantánea. En el caso de la continuación de la historia, la muerte se produce unos días después de haber utilizado la máquina, lo que permite a la persona original darse cuenta de que su réplica *no puede* ser él. En estas circunstancias Quaid podría pensar que su futuro en la Tierra es casi tan malo como morir, pues que haya una réplica suya en Marte no tiene importancia. Sin embargo Parfit argumenta que ser destruido y replicado es casi tan bueno como una vida continuada, pues la réplica tiene todo lo que importa para su supervivencia. La réplica de Quaid continuará con su vida exactamente del mismo modo a como lo haría el propio Quaid si el teletransporte fuese de verdad una forma de viaje. Llevará a cabo los mismos proyectos con idéntica eficacia, cuidará de sus hijos con el mismo amor, e incluso cometerá los mismos errores con igual imprudencia. La réplica de Quaid tiene todo lo que importa para hacer exactamente lo mismo que haría Quaid si siguiera con vida. Todo salvo su identidad.

Parfit defiende que en el caso del teletransporte simple la réplica de Quaid sí que tiene la identidad del original. Su versión del criterio psicológico contempla que una persona A en el tiempo t es la misma que otra persona B en el tiempo t' si existe continuidad psicológica entre ambas independientemente de cuál sea la causa, a no ser que se haya producido una ramificación de forma que en el tiempo t' exista otra persona C en la misma situación, pues la duplicación destruye la identidad personal. De este modo la identidad consiste en el darse una relación de

continuidad psicológica, y además que esa relación sea única –se dé en la forma uno-uno–. Bajo esta perspectiva cabe preguntarse si la unicidad aporta alguna importancia a la identidad personal, o si por el contrario la importancia procede de la continuidad psicológica. Pero dado que la relación entre dos personas no puede cambiar significativamente por el hecho de que exista una tercera, debemos aceptar que lo que de verdad importa es la continuidad psicológica, y la identidad personal obtiene la mayor parte de su importancia del hecho de que normalmente *se da al mismo tiempo* que ésta. La unicidad aún podría representar una leve diferencia, pero comparativamente debería ser insignificante. En consecuencia, siguiendo el criterio psicológico de Parfit, la identidad personal no es lo que importa cuando nos preocupamos por nuestra supervivencia (Parfit, 1984, pp. 469-471).

Volvamos al problema de la duplicación. En el caso del teletransporte, tenemos dos cuerpos cualitativamente idénticos –llamémosles Quaid terrestre y Quaid marciano– que tienen la misma relación de continuidad psicológica con una persona pasada –Quaid original–. A pesar de que ambos cuerpos poseen aquello que nos importa para la supervivencia, la relación no se da en la forma uno-uno, por lo que ninguno de los dos es el Quaid original. Esto es fácil de aceptar para el caso del Quaid marciano, pero de hecho, al contrario de lo que podría parecer, el Quaid terrestre tampoco puede reclamar su identidad, pues aunque existe una continuidad física entre ambos, según el criterio que estamos utilizando la identidad queda definida únicamente por la continuidad psicológica. Como acabamos de ver, al Quaid terrestre no le debería importar no tener la identidad del Quaid original, ya que lo que importa en la supervivencia es la continuidad psicológica y no la identidad. Sin embargo sería ilógico considerar que el Quaid original ha muerto no por algo que le haya sucedido a él, sino por la aparición de una réplica en un planeta distinto. Si eso fuese posible, podría suceder que todos nosotros estuviésemos muertos y no nos hubiésemos enterado, lo cual parece absurdo. En este sentido puede resultar útil modificar nuestro criterio de identidad para que nos permita decir que el Quaid terrestre conserva todavía la identidad del Quaid original, a pesar de que desde el punto de vista del criterio psicológico dicha decisión necesariamente tenga que ser arbitraria.

Al enfrentarnos con el problema de la duplicación en el caso de los dobles de Guy Fawkes dijimos que hacer depender la identidad personal de un hecho trivial no era una posibilidad, pues su arbitrariedad sería incapaz de justificar la importancia que atribuimos a la identidad personal. Sin embargo ya nos hemos visto forzados a admitir que la identidad personal no es lo que importa, por lo que hacerla depender ahora de un

hecho trivial no supone ningún inconveniente. En este sentido Parfit argumenta que si nos encontramos con un caso en el cual según el criterio que estamos utilizando la identidad sea indeterminada siempre es mejor modificar nuestro criterio y ofrecer una respuesta aunque ésta sea arbitraria y no signifique nada, pues hacer que la relación de identidad sea determinada nos permite tener una imagen más ordenada del mundo (Parfit, 1984, p. 436). De este modo los casos de duplicación dejan de ser un problema para convertirse en una evidencia de que la identidad personal no es lo que importa.

Antes de terminar podemos señalar que a pesar de que Parfit circunscribe su razonamiento a la versión amplia del criterio psicológico, el mismo esquema argumentativo es perfectamente aplicable tanto a la versión restringida como al criterio físico, ya que ambos se enfrentan al mismo problema que la versión amplia con los casos de duplicación. De este modo tanto los partidarios del criterio físico como los del psicológico podrían superar el problema de la duplicación recurriendo a una decisión arbitraria si estuviesen dispuestos a aceptar que la identidad personal no es lo que importa. Sin embargo, al igual que sucedía con el cuatridimensionalismo de Lewis y Perry, el precio es tan alto que la mayoría de los que se enfrentaban al problema de la identidad personal se decantaron por otras opciones.

1.6. Conclusiones

Como señalábamos al principio del capítulo, la discusión en torno a los distintos criterios de identidad del PO concluyó con la impresión de que los términos del debate eran estériles. Los partidarios del criterio psicológico respondieron de forma solvente a las dudas planteadas por los problemas de la transitividad y de la circularidad, encontrándose en igualdad de condiciones con los partidarios del criterio físico. Sin embargo ninguno de los dos bandos fue capaz de ofrecer una respuesta satisfactoria a los problemas de la duplicación y de la fusión que no implicase que la identidad personal es una relación indeterminada o que depende de hechos extrínsecos y/o triviales.

De todos modos, a pesar del fracaso del debate entre los criterios físico y psicológico, considero que podemos extraer algunas conclusiones importantes de cara a nuestra investigación. Como ya he señalado, todos los participantes del debate compartían tres creencias. Que el reduccionismo es cierto, que nuestra identidad es determinada y que ésta es precisamente lo que nos importa cuando nos preocupamos por nuestra

supervivencia (*cf.* Ricoeur, 1990, pp. 126-137). Siendo así, era inevitable que la discusión buscara un nuevo marco en el que desarrollarse, pues en los términos en los que estaba planteada se hacía evidente que las tres creencias eran incompatibles. Pero si algo deberíamos aprender del PO es que las tres creencias son incompatibles no debido al planteamiento del problema, sino en todos los casos, por lo que la opción más razonable consiste en abandonar una de ellas. Parfit opta por abandonar la tercera, yo abogaré por el abandono de la primera. El planteamiento narrativo clásico (PNC) parece evitar enfrentarse a esta decisión gracias al enfoque que adopta, completamente distinto al parfitiano y que oculta este problema de base. Sin embargo, como mostraré en el próximo capítulo, es posible interpretar las críticas que los narrativistas hicieron al PO como críticas al reduccionismo, por lo que podemos considerar que su forma de abordar el problema de la identidad personal consiste en abandonar esta creencia para no tener que renunciar a las otras dos, como tendremos ocasión de ver.

Sin embargo, a pesar de haber surgido como una alternativa al debate anterior, al dejar de lado en gran medida dicho debate y la conclusión que podemos aprender de él, la mayoría de las teorías narrativas, tal y como están formuladas, seguirían sin poder enfrentarse a los problemas de duplicación y de fusión de personas, por lo que no parecen aportar soluciones reales al problema parfitiano. Si en un futuro hay dos cuerpos cuya narración les hace ser una evolución de mí mismo, ambos reclamarán mi identidad, lo cual es imposible. Por otro lado si mantenemos un marco materialista no queda claro qué tipo de consideración ontológica tenemos en tanto que producto de una narración. Evidentemente espero poder ofrecer una respuesta a estos problemas, pero en cualquier caso será necesario profundizar primero en el análisis del PNC.

CAPÍTULO 2

EL PLANTEAMIENTO NARRATIVO CLÁSICO

Como decíamos al final del capítulo anterior, el planteamiento narrativo clásico (PNC) se propuso desde un principio como un intento de superar las viejas teorías sobre la identidad personal abandonando casi por completo los presupuestos del planteamiento objetivo (PO). Siendo así no es de extrañar que las primeras aproximaciones al tema surgieran de forma simultánea a los últimos coletazos del debate entre animalistas y neo-lockeanos a partir de autores que no habían tenido ninguna participación en el mismo. En este sentido podemos ver que la obra que se considera fundacional del PNC es el libro *Tras la virtud* de Alasdair MacIntyre (1981), un autor conocido principalmente por sus trabajos sobre filosofía moral.

A partir de la publicación de *Tras la Virtud* la idea de que el Yo podría estar constituido por una narración empezó a calar en el ambiente filosófico,¹ y mientras Derek Parfit respondía con solvencia pero con un

1 No debe entenderse que la obra de MacIntyre es la única causa del auge del narrativismo, sino más bien al contrario. El narrativismo, como una corriente más amplia dentro de las humanidades que interpreta todo el conocimiento humano en términos narrativos y que afecta no sólo a los estudios literarios, sino también a otras áreas de conocimiento como la historia, la antropología, la psicología o por supuesto la filosofía, cobra fuerza en general a principios de los años ochenta (*vid.* Bruner, 1991, p. 5). En este sentido se puede considerar la teoría de MacIntyre como un producto más de esta corriente. De hecho Dennett, otro de los principales defensores del Yo narrativo, publica la primera versión de su teoría el mismo año que MacIntyre de forma completamente

interés cada vez menor las críticas suscitadas entre los partidarios del PO por su *Razones y personas* (*vid.* Parfit, 2004), las publicaciones que constituyen la base del PNC comenzaron a ver la luz. Entre ellas podemos destacar los artículos «Life as Narrative» de Jerome Bruner (1987) y «The Self as a Center of Narrative Gravity» de Daniel Dennett (1988), así como los libros *Las fuentes del yo* de Charles Taylor (1989) y *Sí mismo como otro* de Paul Ricoeur (1990), siendo éste último el que presenta su postura de forma más explícita como una respuesta a la obra de Parfit. En cualquier caso no es mi intención entrar a analizar en este momento las versiones que distintos filósofos han ido presentando sobre el tema de la identidad narrativa. En su lugar me gustaría ofrecer una visión panorámica de las mismas de una forma parecida a la que se ha utilizado para presentar el análisis del debate sobre el problema de la identidad personal desde el PO.

Sin embargo se ha de tener cuenta que, a diferencia de lo que sucedía con el debate expuesto en el capítulo anterior, existen autores relevantes para el desarrollo del PNC que en realidad no se preocupan por el problema de la identidad personal. En este sentido podemos considerar el PNC como una parte de un debate más amplio, el del Yo narrativo, en el cual hay que diferenciar entre los autores que consideran que el Yo está constituido de hecho por narraciones y los que acuden a la narratividad sólo para explicar los mecanismos de autocomprensión con los que contamos. Los primeros aceptan también la tesis de que nos comprendemos a nosotros mismos gracias a nuestros relatos, sin embargo los segundos no aceptan que dichos relatos sean los que nos constituyen. En este sentido, aunque en las siguientes páginas haga referencia a autores de los dos bandos, me interesan especialmente los que defienden que el Yo está constituido por narraciones, pues son los que verdaderamente se enfrentan al problema de la identidad personal y por tanto forman parte del PNC.

Otro aspecto que aumenta la confusión en el debate es que no sólo desde la filosofía se ha venido hablando del Yo narrativo en las últimas décadas, sino que también los psicólogos han desarrollado sus teorías al respecto (*vid.* Rodríguez González, 2003, pp. 124-146). Sin embargo los enfoques de ambas disciplinas son distintos en líneas generales, pues mientras los filósofos cuando hablan del Yo se refieren a una supuesta

independiente (Hofstadter y Dennett, 1981, pp. 348-352). Sin embargo la teoría de Dennett, aunque fue presentada de una forma ya bastante desarrollada en un congreso en Houston en 1983, no empezó a ser conocida hasta la publicación de una versión del texto leído en dicho congreso en 1988, por lo que se considera a MacIntyre en solitario como el precursor de las teorías del Yo narrativo (*vid.* Muñoz Corcuera, 2013).

entidad mental individualizable que según qué autores podría aspirar a la consideración de objeto real, los psicólogos están interesados en el Yo como autoconcepto o como autopercepción, independientemente del tipo de existencia que pudiese tener. En consecuencia, aunque las reflexiones de los psicólogos pueden ser de gran ayuda para entender las reivindicaciones de algunos filósofos y haré uso de ellas cuando sea necesario, serán principalmente estos últimos los que ocupen mi atención en las próximas páginas.

Una de las consecuencias de esta confusión que se produce en el debate sobre el Yo narrativo es la indefinición que existe hasta cierto punto en el PNC. Dado que algunos autores importantes para el desarrollo del mismo en realidad no abordan el problema de la identidad personal es complicado sistematizar sus propuestas, pues no existe un consenso ni siquiera sobre los aspectos a los que debe dar respuesta toda teoría de la identidad narrativa. Siendo así, algunas versiones apenas exploran áreas que son centrales en otras, haciéndose en ocasiones difícil realizar una comparación y saber si se trata de posturas opuestas o complementarias.² Intentaré ofrecer en este capítulo una guía de las preguntas que toda teoría de la identidad narrativa debería responder, pero me centraré únicamente en los puntos mínimos comunes a la mayor parte de las mismas y en algunas críticas realizadas en fechas recientes a la totalidad del planteamiento narrativo clásico, ignorando siempre que no sean pertinentes los aspectos más particulares. Con ello espero mostrar las virtudes y las inconsistencias del PNC y cómo la única forma de solucionar sus problemas es mediante una revisión de sus presupuestos ontológicos y epistemológicos. Para todo esto será de gran importancia el artículo «Against Narrativity» de Galen Strawson, quien realizó el primer gran intento de sistematización de los presupuestos básicos comunes a la mayoría de las teorías narrativas (las dos tesis de la Narratividad³) con la intención de mostrar la debilidad de las mismas.

Considerando lo anterior, el presente capítulo seguirá el siguiente esquema. Primero, mostraré en qué sentido el PNC aborda el problema de la identidad personal tratando de solucionar los errores del PO mediante el abandono del reduccionismo. Segundo, analizaré las críticas que los partidarios de las teorías narrativas hicieron al reduccionismo parfitiano y cómo frente a éste proponen un marco materialista más abierto en el que, mediante las dos tesis de la Narratividad, tengan cabida las dimensiones

2 En este sentido pueden consultarse los trabajos de Dieter Teichert (2004) y Joan McCarthy (2007), que comparan dos teorías de la identidad narrativa bastante diferentes entre sí como son las de Dennett y Ricoeur.

3 La mayúscula es utilizada por Strawson para señalar el carácter específicamente psicológico del término.

fenomenológica y social de nuestra existencia que el reduccionismo ignora. Tercero, expondré algunas críticas que en los últimos años ha recibido el PNC, mostrando que en el fondo revelan la incapacidad de las teorías narrativas para dar una explicación satisfactoria de estas dimensiones fenomenológica y social de nuestra identidad. Por último mostraré el modo en que los narrativistas han intentado responder a las críticas recibidas mediante un nuevo planteamiento narrativo (NPN) señalando los problemas que aún deben de enfrentar, especialmente el hecho de que para desarrollarlo considero que será necesario no sólo abandonar el reduccionismo, sino también el materialismo.

2.1. Entendiendo el giro narrativo

Hasta ahora he venido dando por supuesto que el PNC aborda el mismo problema que abordaban Williams, Shoemaker o Parfit. Esto es, que se preocupa por la identidad numérica de aquello que las personas somos en un sentido más o menos metafísico. Sin embargo, como acabo de señalar, un buen número de teóricos actuales del Yo narrativo parecen no preocuparse ya por el problema de la identidad personal, sino únicamente por los mecanismos que tenemos a nuestra disposición para comprendernos a nosotros mismos (*vid.* Nelson, 2003; Velleman, 2005). Siendo así creo que será útil señalar bajo qué interpretación el PNC es heredero de las posturas neo-lockeanas y por tanto se ocupa del problema de la identidad personal en el sentido que aquí me interesa.⁴

Como señalé en el capítulo anterior, el planteamiento objetivo del problema de la identidad personal partía del intento de sostener tres creencias derivadas de nuestras intuiciones acerca de aquello que son las personas:

1. Que el reduccionismo es cierto.
2. Que la identidad personal es determinada.
3. Que la identidad personal es lo que importa en nuestra supervivencia.

Sin embargo, tal y como pudimos ver, Parfit demostró que resulta imposible sostener las tres creencias al mismo tiempo, pues si aceptamos

4 Podemos señalar aquí que, al igual que el PNC es heredero de la postura neo-lockeana del PO, existen también animalistas en la actualidad, aunque los teóricos de la identidad narrativa apenas se han preocupado de discutir con ellos. Entre sus principales defensores están Eric Olson (1997, 2007) y Paul Snowdon (1990, 2011).

dos cualesquiera de ellas la tercera se revela como patentemente falsa. Al definir el reduccionismo como una simple negación del dualismo cartesiano⁵ Parfit da la sensación de que (1) es una postura que ningún filósofo serio debería poner en duda, por lo que toda su reflexión es alrededor de cuál de las otras dos creencias es preferible mantener, llegando a la conclusión de que para tener una imagen más ordenada del mundo es preferible sostener (2). Según Parfit, cuando no tenemos ningún motivo de peso para decidir acerca de la identidad de una persona es preferible dar una respuesta arbitraria y conservar así la determinación de la identidad, aunque dicha decisión implique que ésta no tenga la importancia que pensábamos. Por el contrario alguien puede considerar que la evidencia de que lo que nos importa en nuestra supervivencia es nuestra identidad es irrenunciable. No obstante, para poder sostener esto nos vemos en la obligación de aceptar que se trata una relación indeterminada, por lo que ante algunas situaciones no podremos decir si un cierto hecho implicaría nuestra muerte o no.

Pero pese a la forma en que Parfit presenta su postura sobre el reduccionismo, su alcance es mayor que el que reconoce explícitamente como veremos a continuación. Tal y como Ricoeur lo expone, al proponer una descripción objetiva e impersonal del mundo el reduccionismo parfitiano deja fuera aspectos cruciales de lo que son las personas, como puede ser la dimensión subjetiva de la experiencia (Ricoeur, 1990, pp. 130-131).⁶ Si leemos bajo este prisma las críticas que no sólo Ricoeur, sino también otros autores como MacIntyre, Taylor o Schechtman hicieron al PO nos damos cuenta que en realidad todas pueden resumirse en dos, ambas referidas a la restrictiva concepción reduccionista que dominaba el enfoque anterior.⁷ No sólo eso, sino que como veremos estos dos tipos de

5 En realidad, como señalé en la primera nota al pie del capítulo anterior, Parfit señala una segunda alternativa al dualismo espiritualista que los no-reduccionistas podrían adoptar, aunque rápidamente trata de demostrar su falsedad. En este sentido, para Parfit el dualismo es el máximo exponente de la postura no-reduccionista (Ricoeur, 1990, p. 129), de tal modo que al oponerse a éste el reduccionismo parece ser la opción más filosóficamente respetable (McDowell, 1997, p. 233).

6 A este respecto es interesante señalar que para otro reduccionista como Shoemaker la postura parfitiana parece implicar también una reducción de los fenómenos mentales a términos físicos o funcionales, acercándose así al eliminativismo o al funcionalismo reductivo (Shoemaker, 1985, p. 139). Véase el apartado «El materialismo frente al problema de la conciencia» del capítulo tres para más información.

7 En este sentido se pronuncia también Manuel Cruz al afirmar que «probablemente constituya un error de enfoque centrar la crítica a la estrategia

críticas desembocan en la defensa de lo que Strawson señaló como las dos tesis de la Narratividad. Con ellas el PNC podía ofrecer en principio una alternativa al PO con la que sostener al mismo tiempo (2) que la identidad personal es una relación determinada y (3) que ésta es lo que nos importa cuando hablamos de nuestra supervivencia sin tener que abandonar el materialismo para recurrir, como Parfit sugería, al dualismo. La idea es la siguiente: siguiendo a Parfit, si no se puede aceptar que la identidad personal sea una relación indeterminada, entonces es mejor dar una respuesta en los casos conflictivos, aunque ésta sea arbitraria (Parfit, 1984, p. 436). Para Parfit cualquier decisión que pudiésemos tomar debía basarse en algún hecho trivial, pero a la luz de las teorías narrativas aparece un hecho que hasta entonces no se había considerado: la propia decisión. Si las personas no están constituidas por su cuerpo, sino por la narración que hacen de su propia vida, entonces el devenir de la narración es lo que determina la identidad de un modo que, en el nuevo marco, no es para nada trivial. La decisión que tomemos, esté basada en el hecho en que esté basada, se convierte por sí misma en un nuevo hecho que posee una importancia vital.

Aunque algunos de los autores que aquí me interesan han adoptado de un modo más o menos implícito esta postura –Ricoeur es casi el único que se enfrenta explícitamente a Parfit, a quien considera «el adversario [...] más temible para [su] tesis» (Ricoeur, 1990, p. 126)–, ninguno de ellos la ha desarrollado por completo, por lo que en realidad no han conseguido ofrecer una defensa contundente de (2) y (3). Así, aunque Marc Slors por ejemplo señala cómo algunos de los experimentos mentales que llevaban a Parfit a defender la indeterminación de la identidad pueden ser solventados si entendemos la historia que contamos sobre nosotros mismos como un tipo de continuidad psicológica, deja sin respuesta los problemas de la fusión y de la fisión, que como vimos fueron los principales responsables del abandono del PO (Slors, 1998, pp. 78-79). Por su parte otros autores, como Dennett, a pesar de defender una postura

parfitiana en su metodología, cuando lo que realmente merece ser sometido a análisis es el fondo del asunto, esto es, su específico reduccionismo» (Cruz, 2005, pp. 40-41). En cualquier caso tampoco es anecdótico el hecho de que la metodología del PO consistiese en el recurso a los polémicos experimentos tecno-científicos a los que se refiere Cruz, pues estos pueden ayudar a despersonalizar las situaciones de un modo sinérgico con el marco reduccionista. Por el contrario, como señala Pedro Enrique García Ruiz, la estrategia del PNC, consistente en la comparación de las personas con personajes literarios, puede utilizarse para resaltar la dimensión subjetiva de nuestra existencia vital (*vid.* García Ruiz, 2009a).

narrativa parecen aceptar que la identidad personal es indeterminada, lo que inevitablemente nos hace pensar en qué motivos podríamos tener entonces para abandonar las teorías de la continuidad psicológica para abrazar la propuesta narrativa abriendo así un nuevo frente de vulnerabilidades que aparentemente no aporta ninguna solución (Dennett, 1991a, pp. 433-434). En cualquier caso, aunque considero que una lectura atenta de las reivindicaciones de los primeros teóricos del Yo narrativo —especialmente MacIntyre, Taylor y Ricoeur— puede ayudarnos a formular una nueva versión de las teorías de la identidad personal que defienda coherentemente (2) y (3), por ahora únicamente detallaré los argumentos que los partidarios del PNC presentaron para atacar al PO y defender las dos tesis de la Narratividad como solución a los problemas detectados.

2.2. Los dos tesis narrativas

A pesar de que, como ya hemos señalado, los autores que se han ocupado del Yo narrativo han ignorado la mayor parte del debate anterior sobre la identidad personal a través del tiempo para ofrecer directamente su propia perspectiva, podemos encontrar entre sus escritos dos tipos de críticas al PO que se pueden interpretar como ataques a la postura reduccionista. A su vez estos dos tipos de crítica se pueden relacionar con la defensa de lo que Strawson denominó las dos tesis de la Narratividad, comunes a la mayor parte de las teorías del PNC y que los narrativistas proponen como una alternativa al reduccionismo. Siendo así, trataré de ofrecer una explicación de las reivindicaciones básicas del narrativismo a raíz de las críticas realizadas al planteamiento objetivo.

El primer tipo de crítica es el ya mencionado, referido a la objetivización de la vida mental que los partidarios del criterio psicológico habían impuesto hasta entonces y que nos hacía considerar los estados mentales como unidades impersonales que podían ser aisladas del resto de la psique para ser eliminadas, alteradas o trasplantadas de una mente a otra sin mayores complicaciones. En respuesta a esta consideración las teorías narrativas plantean una aproximación más subjetiva a nuestra conciencia que tenga en cuenta que nuestras experiencias no pueden ser entendidas de forma independiente al contexto en el que se producen, pues nuestro cerebro organiza nuestra vida mental en forma de narración, dando lugar así a la tesis psicológica de la Narratividad. El segundo tipo de crítica se refiere al hecho de que las consideraciones morales implicadas por la noción de identidad personal nunca tienen sentido fuera del contexto social que las suscita, por lo que al abstraer a la persona de su ambiente las

preguntas por las mismas carecen de sentido. No se puede hablar de responsabilidad moral, por ejemplo, si no se tiene en cuenta al otro ante el que tenemos que responder, y la única forma de hacerlo es mediante una comprensión narrativa de la vida humana, reflexión que cristalizará en la tesis ética de la Narratividad.

2.2.1. La tesis psicológica de la Narratividad

Quizá la mejor forma de entender la crítica relativa a la pérdida que supone la objetivización de las personas al buscar un criterio de identidad es la que Ricoeur expone al hacer su famosa diferenciación entre identidad-*idem* e identidad-*ipse*. Esto es, entre la identidad entendida como respuesta a la pregunta «¿qué?» y la identidad entendida como respuesta a la pregunta «¿quién?» (Ricoeur, 1990, p. 112). Al proponer criterios objetivos, los partidarios del PO se habrían limitado a buscar una respuesta a *qué* es lo que somos, pues la pregunta por el «quién» requeriría necesariamente de una aproximación a nuestra existencia que tuviese en cuenta nuestra subjetividad. De este modo toda la investigación anterior sobre la identidad personal se habría hecho desde el presupuesto de que era posible hacer una descripción *impersonal* de las personas, que era posible decir quién somos después de haber eliminado el «quién». Pues la objetividad supone ignorar, tanto en lo que se refiere a nuestro cuerpo como a nuestra mente, la cualidad de «ser nuestros» que tienen.⁸ El concepto de cuasi-recuerdo, central en la defensa del criterio psicológico, sería el ejemplo máximo de esta despersonalización. El recuerdo que no pertenece a nadie no puede dar cuenta de un «quién», pues es precisamente el «quién» lo que se elimina al *cuasificarlo*. Parfraseando a Ricoeur, los recuerdos propios no pueden ser un caso particular de los recuerdos neutros o cuasi-recuerdos, pues parece ininteligible cómo lo propio puede ser un caso particular de lo impersonal (Ricoeur, 1990, p. 130).

Por eso los criterios objetivos no pueden dar una respuesta satisfactoria al problema, porque lo que nos interesa de la identidad personal es precisamente la pregunta por el «quién». «¿Quién soy?» es una interrogación dirigida a lo que me caracteriza, lo que me hace distinto del

8 A este respecto es importante señalar que la diferenciación entre identidad-*idem* e identidad-*ipse* no se corresponde con la división entre los criterios físico y psicológico. También nuestro cuerpo tiene una dimensión subjetiva, una cualidad de pertenecernos, del mismo modo que los estados mentales, al ponerlos en relación con estados cerebrales, se pueden objetivizar (Ricoeur, 1990, p. 125).

resto, lo que me hace ser la persona que soy. En este sentido lo que nos interesa del problema es la identidad que se esconde detrás de lo que llamamos una «crisis de identidad» (Taylor, 1989, pp. 43-44).⁹ El Yo sólo existe dentro de un marco de valoraciones subjetivas y es ahí donde debe buscarse. Por eso Marya Schechtman ha señalado que el mal planteamiento del debate parfitiano residía en mezclar lo que ella llama las cuestiones de la reidentificación y de la caracterización, pues las preguntas sobre las condiciones lógicas para identificar a una persona a través del tiempo no pueden aportar respuestas a lo que nos preocupa cuando nos preguntamos qué es lo que nos hace ser quienes somos (Schechtman, 1996, pp. 1-2).¹⁰ De una forma todavía más radical lo expresa Mariano Rodríguez González al aseverar que ni la continuidad corporal ni la psicológica podrían ser nunca criterios constitutivos de la identidad personal, sino sólo de evidencia, pues renuncian a lo específico de la condición de persona, esto es, la subjetividad (Rodríguez González, 2003, pp. 96-97).

La búsqueda de una forma de acercarse a nuestra experiencia del mundo que no la traicionase al objetivarla tenía que partir entonces de la introspección, de un análisis de la forma en que subjetivamente percibimos nuestra realidad. Y la respuesta llegó casi al mismo tiempo de dos frentes distintos: MacIntyre desde la filosofía analítica y Ricoeur desde la continental defendieron que experimentamos el mundo de forma narrativa, por lo que si consideramos nuestras narraciones sobre nuestra vida

9 Recientemente Parfit ha señalado que cuando alguien se pregunta por la identidad implicada en una crisis de identidad en realidad ha dejado de preguntarse por la identidad numérica de las personas para pasar a preguntarse por la identidad cualitativa (Parfit, 2012, pp. 5-6). La pregunta sobre quién se es sería sustituible desde su perspectiva por la pregunta sobre qué tipo de persona se es, lo que dejaría de lado lo que de verdad nos preocupa cuando nos preguntamos por nuestra supervivencia. En este sentido argumenta que aunque alguien puede decir que ya no es la misma persona que era antes de casarse eso no significa que casarse haya implicado su muerte. Creo sin embargo que se equivoca, y que la identidad implicada en una crisis de identidad es la misma que nos preocupa al hablar de nuestra supervivencia, pues un cambio brusco en nuestra vida puede llevarnos a interrumpir nuestra narración implicando por tanto nuestra muerte.

10 Patrick Stokes interpreta el abandono de la pregunta por la reidentificación en la teoría de Schechtman como una renuncia a encontrar una solución a los problemas de la identidad numérica, de un modo parecido al que hace Parfit (Stokes, 2012, pp. 86-88). Como en el caso anterior considero que es una lectura equivocada, y que la intención de Schechtman es como sugiero aquí señalar la importancia de atender a nuestra subjetividad.

estaremos teniendo en cuenta la perspectiva subjetiva que hasta entonces se había ignorado.

La tesis de MacIntyre es fácil de entender y se basa en el hecho de que para comprender la conducta de una persona debemos conectarla con toda una serie de factores que la rodean, como pueden ser los motivos que han llevado a la persona a efectuar dicha acción y el contexto en que se realiza (MacIntyre, 1981, pp. 254-255). MacIntyre lo explica a partir de un ejemplo sencillo. Podemos imaginar una situación en la que esté implicada una persona y que pueda ser descrita de las siguientes maneras: «está cavando», «está haciendo ejercicio», «está arreglando el jardín», «se está preparando para el invierno» o «está complaciendo a su mujer». Todas ellas podrían ser descripciones correctas de una misma situación dada, pero sin embargo para poder elegir entre dichas opciones y decidir cómo vamos a describirla es necesario que tomemos en consideración las intenciones de la persona implicada, pues la acción no es la misma si el motivo que ha llevado al hombre al jardín armado con una pala es hacer ejercicio o si es evitar una pelea con su esposa, pues se enmarcan dentro de contextos diferentes. De este modo se nos hace forzoso aceptar que en nuestra mente toda situación está inmersa en una secuencia, una historia que la conecta con todos los aspectos necesarios para su comprensión, y que sin dicha historia la situación resulta ininteligible. Por tanto no es posible individualizar estados mentales del modo en que lo hacían los partidarios del criterio psicológico, sino que se deben considerar secuencias narrativas completas.

Un ejemplo ilustrativo de esta postura podemos encontrarlo en la película *Olvidate de mí* (Michel Gondy, dir., 2004), en la que un neurólogo desarrolla una técnica que permite borrar recuerdos de forma selectiva. Los dos protagonistas de la película deciden someterse al proceso para borrar de su memoria la relación amorosa que han mantenido durante los últimos años. Sin embargo, pese al éxito de la operación, tras ella ambos se encuentran tremendamente desorientados y confusos. La historia de su vida ahora presenta huecos en blanco que desde su propia vivencia personal son significativos. En este sentido, aunque no recuerdan haberse conocido en la playa de Montauk, los dos sienten una vinculación emocional con dicha playa que son incapaces de explicar, lo que les lleva de forma independiente a viajar hasta allí un día en pleno invierno. Inexplicablemente ambos se sienten atraídos nada más verse y vuelven a comenzar una relación que no recuerdan haber tenido. Lo que el neurólogo no había tenido en cuenta es que el recuerdo del día en que se conocieron había resignificado todos sus demás recuerdos relativos a Montauk, por lo que aunque fue capaz de borrar ese día de su memoria, no pudo impedir

que los dos sintiesen que la historia de su vida estaba ahora incompleta y que la respuesta debía encontrarse en la playa. Lo que había ignorado es que nuestra mente se organiza en forma narrativa y que al borrar episodios individuales de las vidas de sus pacientes provocaba que muchos aspectos de las narraciones mediante las que se comprendían a sí mismos se volviesen ininteligibles y demandasen una explicación.

Por su parte la tesis de Ricoeur, que se puede considerar complementaria de la de MacIntyre, procede de su análisis de la temporalidad, y aunque ya había sido adelantada anteriormente (*vid.* Ricoeur, 1980), alcanza su máxima expresión en los tres volúmenes de su magna obra *Tiempo y narración* (1983, 1984, 1985). En ella Ricoeur analiza la experiencia humana de la temporalidad y su relación con la actividad de narrar, afirmando que:

[...] entre la actividad de narrar una historia y el carácter temporal de la existencia humana existe una correlación que no es puramente accidental, sino que presenta la forma de necesidad transcultural. Con otras palabras: el tiempo se hace tiempo humano en la medida en que se articula en un modo narrativo, y la narración alcanza su plena significación cuando se convierte en una condición de la existencia temporal. (Ricoeur, 1983, p. 113)

Esto es así porque, igual que decía MacIntyre, la comprensión de una acción requiere de su disposición en una trama, lo que en última instancia, y para evitar un problema de circularidad en su teoría de los tres momentos de la mimesis,¹¹ implica que tiene que existir una estructura prenarrativa de la experiencia temporal (Ricoeur, 1983, pp. 123-124). Es decir, se hace necesario postular una experiencia temporal originaria, compartida por todos los seres humanos, y que demanda de un desarrollo narrativo (Ricoeur, 1983, pp. 143-144).

La tesis narrativa formulada por MacIntyre y Ricoeur fue acogida de buen grado por los psicólogos cognitivos, entre los que podemos destacar a Bruner, quien en su *Realidad mental y mundos posibles* se apoya en el filósofo francés para defender la existencia de dos modalidades primitivas del pensamiento: la causal –o paradigmática–, que sería la que utilizaríamos para comprender el mundo físico y que daría lugar a la ciencia; y la intencional –o narrativa–, que utilizaríamos para comprender el comportamiento humano (Bruner, 1986, pp. 23-53). De este modo, según Bruner, al explicar nuestras vidas en términos de deseos o creencias

11 Me refiero a lo que Ricoeur denomina el «círculo vicioso de la mimesis», cuyo análisis no resulta necesario en este momento (*vid.* Ricoeur, 1983, pp. 141-146).

estaríamos ya situándonos necesariamente como personajes en una narración y nuestra identidad quedaría definida por tanto en la medida en que somos los personajes protagonistas de una historia determinada: nuestra autobiografía (Bruner, 1987, p. 15).

A lo largo de la década de los ochenta y hasta la actualidad, y debido tanto a la influencia de la obra de estos autores como en general al auge del narrativismo, la creencia de que las personas comprendemos necesariamente nuestra vida de forma narrativa se ha convertido casi en un tópico que no necesita mayor demostración, dando lugar a lo que Strawson ha denominado la tesis psicológica de la Narratividad, y que se puede formular como la creencia de que «uno ve, vive o experimenta su vida como una narración o una historia de algún tipo, o al menos como un conjunto de historias» (Strawson, 2004, p. 189)¹². Se trata de una tesis descriptiva acerca de cómo es la experiencia humana y su alcance se pretende universal. Independientemente de lo que en última instancia sean las personas, acerca de lo cual no tiene nada que decir, la tesis psicológica de la Narratividad defiende que éstas se aprehenden e interpretan a sí mismas en términos narrativos. En este sentido Dennett –para quien el Yo no es más que una ficción fruto de nuestra propia autointerpretación– compara la necesidad humana de contar una historia de nuestra vida con otros comportamientos que se dan en el reino animal de forma automática como la construcción de presas por parte de los castores o la elaboración de telas por parte de las arañas:

*Nuestra táctica fundamental de autoprotección, de autocontrol y de autodefinición no consiste en tejer una tela o construir una presa, sino en contar historias, y más particularmente, en urdir y controlar la historia que contamos a los demás –y a nosotros mismos– sobre quiénes somos. Y al igual que las arañas no tienen que pensar, consciente y deliberadamente, en cómo deben tejer sus telas, y al igual que los castores, a diferencia de los *profesionales* humanos de la ingeniería, no proyectan consciente y deliberadamente las estructuras que construyen, nosotros (a diferencia de los *profesionales* humanos del contar historias) no imaginamos consciente y deliberadamente qué narraciones contar ni cómo contarlas. (Dennett, 1991a, p. 428)*

La tesis psicológica de la Narratividad es la reivindicación básica de todas las teorías del Yo narrativo, y sostenerla es suficiente para ser considerado narrativista. En cualquier caso es importante señalar que, evidentemente, tal y como está formulada deja un margen muy amplio a qué grado de estructuración implica, pues mientras la mayoría de los

12 La traducción es mía. El original es: «one sees or lives or experiences one's life as a narrative or story of some sort, or at least as a collection of stories.»

autores que hemos venido citando hasta ahora defienden que toda nuestra vida debe formar parte de una misma narración, otros como David Velleman sostienen la existencia de infinidad de narraciones mínimas desconectadas entre sí (Velleman, 2005, pp. 222-223). Podemos ver así la gran variabilidad que existe entre las distintas teorías ya desde su nivel más básico.

2.2.2. La tesis ética de la Narratividad

La segunda crítica que los partidarios del PNC lanzaron contra el PO parte de la importancia que el problema de la identidad personal tiene para lo que Schechtman denomina las cuatro características básicas de la existencia de las personas: la supervivencia, el interés por nosotros mismos, la compensación futura y, muy especialmente, la responsabilidad moral (Schechtman, 1996, p. 2). En efecto parece claro que para poder explicar por qué y bajo qué condiciones una persona es responsable de lo que hizo en el pasado a pesar de los cambios que haya podido experimentar desde entonces debemos poder responder primero a la pregunta por la identidad de dicha persona, pues uno sólo puede ser responsable de sus propias acciones. Del mismo modo, sólo si somos capaces de considerar la vida de una persona en su conjunto podremos decir si dicha persona posee una orientación al bien, pues un juicio de ese tipo no puede realizarse en base a una única acción tomada individualmente. Siendo así, resulta necesario que podamos contar con un criterio que considere las vidas de las personas como una unidad con una duración temporal extensa y que al mismo tiempo garantice que los juicios de identidad siempre serán del tipo todo o nada (*cfr.* Schechtman, 1996, p. 149). Por otro lado, aunque fuese posible elaborar un criterio que cumpliera con estos requisitos a partir del PO, cosa que ya hemos visto que no lo es, además debemos tener en cuenta que las cuestiones de responsabilidad moral sólo tienen sentido en el espacio social que las suscita, por lo que nuestro criterio de identidad debe considerar también que la vida de las personas se da siempre en comunidad. En este sentido Cruz señala que la importancia del concepto de identidad personal se debe a nuestra necesaria convivencia con otras personas, ya que «la sociedad ni tan siquiera puede funcionar sin el ineludible juego de adscripciones e imputaciones, reconocimientos y censuras, que vértebra las relaciones interpersonales» (Cruz, 2005, pp. 48-49). Utilizando la formulación todavía más extrema de MacIntyre, las personas debemos responder a la

imputación de identidad estricta porque «soy para siempre cualquier cosa que haya sido en cualquier momento para los demás, y puedo en cualquier momento ser llamado para responder de ello, no importando cuánto pueda haber cambiado» (MacIntyre, 1981, p. 267). Es esta vertiente social de nuestra existencia la que demanda y hace que nuestra identidad sea determinada. En consecuencia la continuidad física o psicológica nunca podrá dar cuenta de este hecho, pues la unidad de la vida de las personas no depende únicamente de sí mismas.¹³ Siguiendo a Ricoeur, sólo si entendemos a las personas como a los personajes de una narración podemos darles el tipo de unidad requerida:

¿Cómo, en efecto, un sujeto de acción podría dar a su propia vida, considerada globalmente, una cualificación ética, si esta vida no fuera reunida, y cómo lo sería si no en forma de relato? (Ricoeur, 1990, p. 160)

Pero una vez que hemos aceptado la tesis psicológica de la Narratividad el problema al que nos enfrentamos entonces no es si las personas experimentan su vida en forma de narración, ya que esto se ha presupuesto como una necesidad, sino qué condiciones debe cumplir dicha narración para poder dar forma a una vida buena y constituir nuestra identidad. El primer requisito, ya lo hemos avanzado, es que la vida de una persona tiene que ser reunida en una única narración, por lo que se debe imponer una estructura mucho más rígida que la demandada por la tesis psicológica de la Narratividad. El segundo requisito es que la narración debe estar abierta de algún modo para dar cabida a la dimensión social de la persona. En realidad, como hemos visto, el primer requisito está supeditado al segundo, pues la necesidad de que toda nuestra vida sea reunida en un relato procede de la imputación de identidad estricta que demanda nuestra existencia social. Sin embargo esto no ha sido señalado con la intensidad apropiada anteriormente, lo que ha derivado a veces en una mala comprensión del enfoque que se debía dar al problema. En cualquier caso volveré sobre este aspecto más adelante. De momento me

13 Para José Óscar Benito Vicente éste precisamente parece ser el principal problema del PO. Para ejemplificar su postura señala que una réplica de Julio César que estuviese pensando al mismo tiempo que el emperador romano en la conquista de un territorio se diferenciaría de él en un aspecto fundamental: sólo el Julio César original tendría la capacidad efectiva de conquistar ese territorio gracias al rol social que desempeña y que le otorga el mando de los ejércitos de Roma. El PO no puede ofrecernos una solución al problema de la identidad personal porque ignora este aspecto social de nuestra existencia (Benito Vicente, 2003, pp. 81-83).

centraré en decir un poco más de cada uno de los dos requisitos tal y como han sido entendidos hasta el momento por los partidarios del PNC.

En cuanto al primer requisito, el que nuestra vida tenga que ser reunida en una única narración, trae consigo un problema, y es el de la inteligibilidad de dicho relato. Podría ser que nuestra vida fuese tan azarosa que fuésemos incapaces de formar una narración comprensible de la misma. Esto es precisamente lo que ocurre cuando decimos que nuestra vida carece de sentido (MacIntyre, 1981, p. 268) o que sufrimos una crisis de identidad (Taylor, 1989, pp. 43-44). En un caso así lo que sucede es que somos incapaces de valorar quiénes somos y qué es importante para nosotros porque nuestra narración es incomprensible debido a que no tenemos un objetivo, una meta para nuestra vida, lo cual es una condición necesaria para hacer inteligible una narración.¹⁴ Por eso la forma de otorgarle una unidad narrativa a nuestra existencia tiene que consistir en dirigirla hacia un clímax o *telos*. Una hoja de ruta que al señalar qué es lo que deseamos conseguir nos permita valorar las situaciones en las que nos encontramos y actuar de la mejor forma posible para conseguir nuestro objetivo. Sólo así nuestra vida cobrará un sentido que, evidentemente, será únicamente comprensible en tanto clímax de la narración. Y qué mejor forma de dirigir nuestra vida que hacerlo hacia aquello que es bueno para nosotros, adquiriendo de este modo nuestra existencia la forma de una búsqueda del bien (Taylor, 1989, p. 68).

Sin embargo, y aquí entra en juego el segundo requisito, aunque el Yo esté constituido en parte por su propia auto-interpretación narrativa, dicha interpretación está realizada utilizando un lenguaje determinado. Y dado que no puede existir un lenguaje privado, tampoco tiene sentido la idea de un Yo aislado, pues uno sólo es un Yo entre otros Yoes constituidos en el mismo contexto lingüístico (Taylor, 1989, p. 51). En consecuencia el lenguaje en el que está articulado el Yo lleva ya implícitas las valoraciones morales propias de la sociedad en la que se inscribe, abriéndose así el camino para la introducción de la dimensión social de la persona en las

14 MacIntyre lo explica mediante un símil con dos obras de Franz Kafka. En *El castillo* y *El proceso* el personaje de K se encuentra repentinamente envuelto en narraciones que le resultan ininteligibles, lo que le imposibilita dar sentido a su vida. Yendo un paso más allá, podemos ver que en *El proceso* los agentes que se presentan en la casa de K al principio de la novela le dicen que se encuentra «detenido», implicando con ello que la narración de su vida no avanza hacia ningún lugar. En este sentido es comprensible que Kafka no fuese capaz de terminar de escribir sus novelas, pues la noción de desenlace sólo tiene sentido en una narración inteligible (MacIntyre, 1981, p. 263).

perspectivas narrativas.¹⁵ Camino que puede ser completado de la mano de MacIntyre, quien da el paso desde el nivel lingüístico al narrativo.

Señala MacIntyre que entramos a formar parte de una narración desde el mismo momento en que nacemos. Sin embargo no lo hacemos en un escenario que hayamos elegido, ni la acción en la que participamos es de nuestra autoría, pues no es en nuestra propia narración en la que empezamos a existir, sino en las narraciones de otros. Nosotros no aprenderemos a hablar hasta muchos meses después de haber nacido, sin embargo sería absurdo considerar que no existimos antes de eso. Nuestros padres nos dan un nombre y comienzan a contar historias sobre nosotros, a interpretar nuestro comportamiento y la forma en que afecta a sus vidas desde el mismo momento en que nacemos. Desde esta perspectiva nuestro propio drama personal se encaja en otros dramas de los que también formamos parte, aunque nuestra participación en ellos puede ser mínima. Así, aunque somos los protagonistas de nuestra historia, en la de nuestros padres somos sólo un personaje secundario, y en la del médico que nos ayudó a nacer apenas un figurante. Sin embargo, aunque nuestro papel pueda ser menor en las narraciones de otros, no por ello nos doblegamos al guión que los demás nos han escrito en sus vidas, sino que podemos intervenir en ellas a voluntad. Del mismo modo, los demás personajes de nuestro propio drama escriben ellos mismos sus líneas, por lo que debemos admitir que no somos más que meros coautores de nuestra historia (MacIntyre, 1981, p. 263).¹⁶

En este sentido podemos ver que todos ocupamos un determinado sitio en el entramado de historias que forman nuestra sociedad, un papel al que debemos enfrentarnos, pues somos necesariamente hijos de tal padre, ciudadanos de tal país, miembros de tal grupo, etc., y como tales heredamos una serie de deberes, obligaciones y expectativas sobre nuestra vida que nos confieren nuestra propia particularidad moral. Y aunque nuestra vida pueda derivar hacia posiciones distintas a las que poseemos al nacer, siempre tendremos algún tipo de papel social, y la vida buena para

15 Una alternativa distinta es la propuesta por Ricoeur, para quien la dimensión social de la persona entra de mano de la identidad-*ipse*, cuyo ejemplo paradigmático es la fidelidad a la palabra dada. Según su teoría la identidad narrativa llevaría consigo la obligación de mantenerse fiel a uno mismo de tal manera que los otros puedan contar con que cumpliremos nuestras promesas (Ricoeur, 1990, p. 168). Es necesario estar disponible para los otros para permanecer siendo la misma persona.

16 En realidad la mayoría de los partidarios del PNC han ignorado esta idea de MacIntyre que considero fundamental. Quizá porque crea un nuevo problema: el de cómo se relaciona nuestra narración con las narraciones de los demás. Volveré sobre ello cuando desarrolle el NPN en el capítulo quinto.

nosotros debe ser una vida buena para cualquiera que pueda ocupar nuestro lugar, por lo que la narración de nuestras vidas siempre estará afectada por el resto de narraciones de las que formamos parte (MacIntyre, 1981, p. 271).

En este momento es importante señalar que dado que la tesis psicológica de la Narratividad no implica la necesidad de tener una narración tan estructurada de la propia vida como la que estamos exponiendo siempre es posible que nuestra narración no cumpla con estos requisitos. Podemos recordar por ejemplo sin ninguna extrañeza al protagonista de la novela de Max Frisch *No soy Stiller*, quien tras huir de su país e intentar suicidarse se niega a aceptar tanto el nombre como las responsabilidades de su vida anterior. En un caso así, y dado que como hemos visto nuestra narrativa se encuentra afectada por el resto de narraciones en las que participamos, la sociedad puede intentar forzar a dicha persona a reconocer su identidad, como sucede en el caso del protagonista de la novela, a quien el estado de Suiza lleva a juicio para obligarle a aceptar sus responsabilidades legales. Si esta fuera la situación, y como sucede en la obra la persona implicada finalmente reconociese su responsabilidad, únicamente tendríamos ante nosotros a una persona poco ética que debería reestructurar su vida. Sin embargo puede ocurrir que alguien con un severo trastorno psicótico se niegue a aceptar el papel que el resto de la sociedad le impone, o que directamente sea incapaz de ofrecer una narración mínimamente estructurada de su vida al completo. Considerando este tipo de situaciones en las que debido a una enfermedad mental una persona es incapaz de ofrecer un relato único y cohesionado de toda su vida Schechtman señala que aunque no es necesario que todos los seres humanos tengan una perspectiva narrativa sobre sus vidas con los requisitos que hemos expuesto, sí es necesario que la tengan para alcanzar plenamente la condición de persona, por lo que en última instancia no todos los seres humanos serían personas (Schechtman, 1996, p. 119).

Este tipo de consideraciones dan lugar a lo que Strawson denomina la tesis ética de la Narratividad, y que puede ser formulada como la afirmación de que «una perspectiva Narrativa rica es esencial para tener una vida buena, para alcanzar verdadera o completamente la condición de persona» (Strawson, 2004, p. 189)¹⁷. Como se puede ver, se trata de una tesis normativa, no descriptiva, acerca de cómo deberían las personas vivir su vida. Pero a pesar de ello, de que se trata de una creencia que no es necesario sostener en toda teoría del Yo narrativo, y de que algunos autores como Dennett no se acercan de un modo tan explícito al tema de la

17 La traducción es mía. El original es: «a richly Narrative outlook on one's life is essential to living well, to true or full personhood.»

ética¹⁸, es una tesis que se encuentra ampliamente aceptada, quizá debido a que es la base principal de las posturas de tres de los primeros y principales defensores del Yo narrativo como son MacIntyre, Taylor y Ricoeur. Por otro lado, tal y como está formulada, aceptarla es un requisito para sostener una teoría de la identidad personal de tipo narrativo (recordemos que no todos los narrativistas aceptan el PNC), pues implica la apreciación de que no simplemente nos comprendemos a nosotros mismos de forma narrativa, sino que al hacerlo constituimos nuestra identidad.¹⁹ En cualquier caso la mayoría de los autores han olvidado que lo realmente importante de la crítica al planteamiento parfitiano que da lugar a la defensa de la tesis ética de la Narratividad es la necesidad de dar cabida a la dimensión social de nuestra existencia, y no tanto la importancia de tener una narración altamente estructurada y orientada hacia un objetivo definido. Esto último era una forma de alcanzar lo primero, no un fin en sí mismo.

2.3. Críticas al narrativismo

Como se puede ver a partir de lo que acabamos de decir, las teorías del Yo narrativo sólo tienen en común la creencia de que nuestra vida mental, para ser inteligible, tiene que adoptar la forma de uno o varios relatos (esto es, comparten la tesis psicológica de la Narratividad). Sin embargo casi todas coinciden también en otros tres aspectos básicos que aunque no se han desarrollado explícitamente se desprenden de lo anterior: (1) conciben el Yo no como un simple sujeto de experiencias, sino como un agente

18 Aunque si analizamos más profundidad su postura veremos que en el fondo también apoya una versión de la tesis ética de la Narratividad. Para ello véase mi «Problemas y aciertos de la teoría del Yo narrativo de Dennett: Aportaciones al debate sobre la identidad personal» (Muñoz Corcuera, 2013).

19 En una revisión posterior de su teoría, Strawson separó acertadamente la tesis ética de la Narratividad de lo que llama la tesis de la auto-constitución narrativa. De este modo la primera sigue siendo una tesis normativa que defiende que comprendernos a nosotros mismos de forma narrativa es necesario para tener una vida buena, incluso esencial para ser una persona. La segunda por su parte es una tesis descriptiva que asegura que al hacerlo además constituimos nuestra identidad (*vid.* Strawson, 2012). Sin embargo en este momento la diferenciación no es importante, pues al considerar el PNC en tanto planteamiento del problema de la identidad personal estoy dando por supuesta también la tesis de la auto-constitución narrativa, por lo que consideraré la tesis ética de la Narratividad en su forma original que combina ambas apreciaciones.

capaz de actuar y tomar decisiones que (2) está encarnado en un cuerpo determinado del que se no se puede separar²⁰ y que (3) está embebido en un contexto social (*cf.*: Schechtman, 2011, pp. 401-407). Defender estas creencias es una condición suficiente para cualquier teoría del Yo narrativo. En cualquier caso, existe una amplia serie de variantes que afectan a la manera en que dichas creencias pueden insertarse en teorías más desarrolladas. Así, como dijimos al principio del capítulo, algunas versiones defienden que el Yo está constituido de hecho por narraciones mientras que otras sólo dicen que hace uso de ellas para comprenderse; unas aseguran que el Yo es una entidad real mientras otras sostienen que sólo es una ficción; y finalmente unas defienden que es necesario que toda la vida de un Yo esté unificada en una narración altamente estructurada para poder hablar de cuestiones morales mientras que otras ignoran por completo dicho aspecto o les resulta suficiente con la existencia de múltiples narraciones desconectadas entre sí. Quizá la teoría narrativa tipo –y la que me interesa como respuesta al problema de la identidad personal– sería una que se decantase por la primera opción en todas estas disyunciones, asegurando que el Yo es una entidad real²¹ constituida totalmente por una única narración altamente estructurada que es de gran relevancia para las cuestiones morales. Sin embargo prácticamente cualquier combinación ha sido defendida en algún momento.

Pero independientemente de la existencia de variantes dentro de las teorías narrativas del Yo, en los últimos años han aparecido diversas críticas a la totalidad del PNC que plantean serias dificultades para su viabilidad sea cual sea la forma que adopte. Entre ellas merece la pena destacar el ya mencionado artículo de Strawson «Against Narrativity», publicado en 2004 y que se puede considerar el inicio de esta corriente crítica. El argumento de Strawson se dirigía tanto a la validez de la tesis psicológica de la Narratividad, a la que considera falsa en cualquier

20 Dennett otra vez sería una excepción (Dennett, 1991a, pp. 440-441).

21 Señala Kristjánsson que en realidad la mayoría de los partidarios de las teorías narrativas han ignorado el problema de la realidad de los Yoes al no abordarlo explícitamente, siendo muchas veces difícil saber cuál sería la postura de algunos de ellos. Esto podría tener su origen en la ambigua posición de MacIntyre al respecto, y que puede generar la falsa impresión de que se trata de un problema superado. Sin embargo existen tanto realistas como anti-realistas, por lo que se trata de una distinción importante dentro de las posturas narrativas (Kristjánsson, 2009, pp. 181-182). Por mi parte tomo a la mayoría de los defensores del narrativismo por realistas, pues parecen otorgarle cierto poder causal y le atribuyen una importancia en cuestiones morales que difícilmente se le podrían atribuir a una entidad ficticia, aunque existen diversidad de opiniones al respecto (*cf.*: Stokes, 2012).

sentido que no sea trivial, como a la de la tesis ética de la Narratividad, a la que considera no sólo equivocada sino también potencialmente perjudicial para la vida de algunas personas que el denomina «Episódicas» y que tendrían una experiencia esencialmente no-narrativa de sus vidas. La importancia de esta crítica es evidente, pues si los dos pilares sobre los que se sustentan las teorías narrativas resultasen ser insostenibles el problema de la identidad personal retrocedería de golpe hasta el PO, pues los narrativistas se habrían mostrado incapaces de solucionar los mismos dos problemas que achacaban a Parfit. Es decir, el PNC se habría mostrado incapaz de ofrecer una alternativa al reduccionismo que reconociese las dimensiones social y subjetiva de las personas. De este modo, si las dos tesis de la Narratividad fuesen falsas, seguiríamos teniendo que enfrentarnos al dilema planteado por Parfit: las tres creencias que tradicionalmente se han vinculado a la idea de identidad personal son incompatibles entre sí, por lo que debemos abandonar una de ellas. Y si nos decantamos por abandonar el reduccionismo como hizo el PNC, entonces tendremos que buscar un nuevo marco ontológico. Por otro lado, si alguna versión del narrativismo pudiese todavía subsistir, sin duda debería hacerlo adoptando una forma completamente nueva, dado que las críticas de Strawson se dirigen a la misma base del proyecto narrativo.

En cualquier caso, el de Strawson no ha sido el único argumento que se ha esgrimido contra las teorías narrativas, pues podemos encontrar por lo menos otro tipo de crítica que también señala la imposibilidad de dar cabida a la dimensión subjetiva de la experiencia en el PNC. Este segundo tipo de crítica se refiere al hecho de que si estamos constituidos por una narración, no queda claro quién es el sujeto que experimenta conscientemente los sucesos que en ella se contienen, pues no parece posible que sea la propia narración la que experimente dolor, por ejemplo (Menary, 2008, p. 72). Ante esta dificultad las posturas narrativas necesitan o bien defender que el Yo es una ficción, con lo que dejan de tener en consideración nuestras experiencias conscientes, o bien admitir que existe un Yo fenoménico previo a todas las narraciones y que es el sujeto de las mismas (*vid.* Carr, 1998; Zahavi, 2007). Este disyuntiva abrirá el camino a lo que denomino el nuevo planteamiento narrativo (NPN), que trata de dar cabida a la dimensión subjetiva de la experiencia y a nuestra vertiente social diferenciando entre un Yo fenoménico y un Yo narrativo extendido, pero será mejor no adelantarnos a nuestros propios pasos.

Un tercer tipo de crítica que ha recibido el PNC se refiere a la ambigüedad con la que se ha utilizado el término «narración», lo que ha llevado a comparar con una libertad excesiva nuestra vida mental con una

narración literaria o por el contrario a utilizarlo en un sentido tan débil que sus implicaciones son triviales (*vid.* Christman, 2004; Lamarque, 2004, 2007). Sin embargo, a pesar de la importancia de esta crítica, no le prestaré atención en este momento, pues no contribuye a la argumentación que intento presentar en este capítulo. Volveré sobre ella en el capítulo seis una vez haya demostrado que para solucionar el problema de la identidad personal es necesario aceptar un cambio en nuestros presupuestos ontológicos abandonando el materialismo.²²

Un último aspecto que me gustaría señalar antes de empezar es que en este apartado únicamente expondré las dos críticas que afectan al hecho de que las distintas teorías que se engloban en el PNC en realidad no pueden explicar las dimensiones subjetiva y social de nuestra existencia tal y como han sido formuladas por sus autores. Cuando lo considere oportuno indicaré algunas respuestas que han recibido estos argumentos por parte de otros defensores del PNC, pero dejaré su refutación para el momento en que exponga la alternativa que constituye el NPN.

2.3.1. *La falsedad de las dos tesis de la Narratividad*

La crítica lanzada por Strawson contra el PNC trata de demostrar que las experiencias de las personas de forma natural nunca son más que trivialmente narrativas en el sentido en el que preparar café implica una narración porque cada uno de los pasos necesarios para hacerlo requiere de una visión global del proceso para cobrar sentido (Strawson, 2004, p. 198), lo cual es insuficiente –o irrelevante– para hacer que las reivindicaciones de los narrativistas sean la base para hablar de aquello que debe

22 En el análisis que hace Fernando Broncano de las críticas que ha recibido el PNC se distinguen tres problemas que las teorías narrativas deberían afrontar. El problema del escepticismo (que se refiere al hecho de que si la identidad se constituye mediante una narración, siempre es posible que ésta incurra en una fabulación de los hechos), el problema del diacronismo (en referencia a la crítica de Strawson) y el problema del egocentrismo (si nuestra identidad se constituye mediante nuestro propio relato, las narraciones que los demás hacen sobre nosotros parecen no tener ningún papel) (Broncano, 2013, pp. 187-188). El problema del diacronismo será abordado en este capítulo y en el quinto, mientras que el problema del escepticismo lo considero lo suficientemente relacionado con el problema del concepto de narración que se esté usando que lo trataré junto a éste en el capítulo seis. Por su parte el problema del egocentrismo ya fue señalado más arriba en la nota al pie 16, y será abordado en el capítulo quinto cuando desarrolle el NPN.

importarnos cuando nos preocupamos por nuestra supervivencia²³ o por la responsabilidad moral. Siendo así debemos tener en cuenta que aunque algunas personas puedan sentirse cómodas dando a sus vidas una estructura narrativa más compleja –como reivindican los teóricos del PNC– y estar comprometidas a vivirlas de acuerdo a una narración en un sentido del término lo más significativo posible, para muchas otras no sólo resulta anti-natural hacerlo, sino también perjudicial.

La reivindicación de Strawson parte de una diferenciación entre dos formas diferentes en que las personas experimentan su existencia en el tiempo que asume que tienen bases genéticas y que dan lugar por tanto a dos tipos de personas: Diacrónicas y Episódicas (Strawson, 2004, pp. 190-191). Las personas Diacrónicas tienden a verse a sí mismas en tanto Yo es como entidades que existían en un pasado lejano y que seguirán existiendo en el más remoto futuro, mientras que las Episódicas por el contrario carecen de dicha sensación. Aunque un Episódico puede recordar hechos pasados, y conceder que son experiencias que fueron tenidas por él en tanto ser humano, no siente que dichas experiencias le pertenezcan en tanto Yo, pues no tiene ningún sentimiento de propiedad hacia las mismas. Las siente ajenas hasta tal punto que considera que fueron experimentadas por un Yo distinto. Podría interpretarse esta sensación como un fallo en la auto-percepción de dicha persona, como los Diacrónicos tienden a hacer cuando se enfrentan a un Episódico. Pero con el mismo grado de autoridad podemos suponer que se trata del registro de un hecho acerca de lo que un Yo concreto es, pues no tenemos motivos para dudar de la fiabilidad de sus sensaciones subjetivas.

La pregunta entonces es por la relación que existe entre la Diacronicidad y la Episodicidad, en tanto modos de experimentar nuestra existencia temporal, y la Narratividad en tanto manera de comprender nuestra vida, donde Narratividad se debe entender en un sentido no trivial del término. Y la respuesta es que para ser Narrativo es necesario no sólo ser Diacrónico, pues uno puede concebir su vida como una unidad y dar un registro secuencial de sus acciones gracias al sentido en que experimentamos nuestra vida de forma narrativa en un sentido trivial, sino que también hace falta tener una tendencia a buscar en la propia vida una estructura formal y que ésta sea del tipo que tienen las historias de corte clásico (Strawson, 2004, pp. 200-201). Siendo así debemos aceptar que la

23 De hecho para Strawson no deberíamos preocuparnos nunca por nuestra supervivencia. Pero en cualquier caso, si debiésemos hacerlo, esta preocupación no podría justificarse en base a un sentido trivial de Narratividad (*vid.* Strawson, 2007).

tesis psicológica de la Narratividad es falsa, pues ni las personas Episódicas ni algunas Diacrónicas experimentan su vida en forma de narración en ningún sentido que no sea trivial.²⁴

Pero aunque no sea cierto que todas las personas experimenten su vida como una narración de forma natural aún podría argumentarse que debe tomarse como un deber moral el intentarlo, pues según la tesis ética de la Narratividad sólo así podemos alcanzar una vida plena y desarrollarnos como personas. Sin embargo una vida Episódica puede ser igual de feliz y satisfactoria que una Diacrónica, incluso en mayor medida. No sólo eso, sino que las cualidades morales no se ven afectadas por nuestra experiencia temporal. La idea que late de fondo en la tesis ética de la Narratividad, según Strawson, es que es necesario tener una comprensión profunda de uno mismo para poder actuar de una forma ética, y que ésta auto-comprensión sólo puede darse de forma Narrativa. Sin embargo aunque se puede conceder que es necesario un cierto grado de auto-comprensión para tener una vida buena, ésta puede darse sin ningún tipo de Narratividad. Uno puede profundizar en su propia vida sin tener una vinculación explícita con su pasado, igual que los músicos pueden mejorar con su instrumento sin recordar las largas horas de práctica (Strawson, 2004, p. 205). Del mismo modo uno puede ser un buen amigo para alguien sin importar el valor que le dé a las experiencias compartidas en el pasado, pues la amistad se demuestra en cómo se es en el presente (Strawson, 2004, p. 207). Si vamos todavía un poco más lejos podemos ver además que no sólo no es necesario tener una comprensión Narrativa de la propia vida para tener una vida plena, sino que hacerlo puede resultar perjudicial para algunas personas, que sienten que considerarse a sí mismas como si fuesen los personajes de una Narración supone deshumanizarse y falsear su propia vida (Strawson, 2004, p. 205; Vice, 2003, p. 105). Aunque los hechos se narren de un modo correcto, imponerles una estructura Narrativa es sentido por algunos como un alejamiento de la verdad, una pérdida de la perspectiva correcta bajo la que se debe entender su vida. Esta sensación de alienación y enajenación puede provocar finalmente graves problemas emocionales, por lo que adoptar una perspectiva Narrativa, lejos de desembocar en un mejoramiento de la vida, puede suponer para algunas

24 En realidad Strawson admite que el caso límite de Narratividad ni siquiera implica ser Diacrónico, pues una persona Episódica puede sentir la necesidad de dar una unidad narrativa a su vida precisamente por su falta de sensación de tener una existencia temporal dilatada. En cualquier caso para simplificar podemos asumir que en principio para ser Narrativo hay que ser Diacrónico, pues de forma general son dos características que se dan juntas.

personas un grave daño psicológico, demostrándose también así la falsedad de la tesis ética de la Narratividad.²⁵

El artículo de Strawson ha tenido un amplio impacto en las concepciones narrativas del Yo, y varios de los defensores de la Narratividad han intentado dar respuesta a sus argumentos. Entre ellos podemos destacar por el momento a Anthony Rudd, quien basa su defensa de la tesis psicológica de la Narratividad en que la narrativa implicada en el hecho de preparar café no es en absoluto trivial, pues existen concepciones alternativas del modo en que deben comprenderse las acciones humanas (Rudd, 2007, p. 63), mientras que asegura que la tesis ética de la Narratividad no se ve afectada por el argumento de Strawson porque para poder hablar de amistad o de cualquier otro tipo de virtud no basta con considerar una amabilidad de tipo Episódico, sino que es necesario tener en cuenta que la noción implica una cierta temporalidad que sólo puede ser valorada de forma Diacrónica (Rudd, 2007, p. 70). También Stokes ha intentado responder a Strawson atacando su postura neutral con respecto a los caracteres Episódico y Diacrónico, especialmente porque no ofrece una fenomenología de la experiencia Diacrónica del tiempo y sin ella no puede justificar que ambas posturas sean moral y epistemológicamente válidas en el mismo grado. Si intentase hacerlo debería aceptar que la experiencia Diacrónica es preferible a la Episódica, por lo que su argumento perdería fuerza (Stokes, 2010, pp. 120-121).

Sin embargo, pese a que las respuestas de Rudd y Stokes tienen cierto peso —de hecho volveré más adelante sobre la respuesta de Rudd a la acusación de trivialidad—, no parece que sean capaces de eliminar el efecto de la crítica de Strawson, que deja al PNC al borde de la muerte al poner de manifiesto su incapacidad para resolver las dos carencias detectadas en el PO. Y es que, como Schechtman señaló acertadamente, lo que el ataque de Strawson revela es una tensión subyacente en el PNC entre las exigencias que demanda el atender simultáneamente a las dimensiones subjetiva y social de nuestra existencia (Schechtman, 2007, pp. 168-169). Si intentamos que nuestra narración se adapte lo más posible a nuestra experiencia subjetiva del mundo acabamos teniendo que renunciar a toda

25 Como señalaba anteriormente, en una revisión de su teoría Strawson separa la tesis ética de la Narratividad de la tesis de la auto-constitución narrativa, ofreciendo argumentos extra en contra de esta última. En este sentido señala que afirmar que nos constituimos a partir de lo que pensamos sobre nosotros mismos implica la idea de que tenemos un amplio acceso a aquello que somos, lo que se puede cuestionar fácilmente a partir de la existencia del subconsciente (Strawson, 2012, p. 87). En cualquier caso esta crítica no se aplicará a mi versión revisada del NPN, por lo que no le prestaré más atención.

perspectiva narrativa, al menos en cualquier sentido significativo del término que nos permita tener en cuenta al mismo tiempo nuestra vida social. Si por el contrario intentamos que nuestra narración tenga las características apropiadas para dar cuenta de la dimensión social de la identidad personal, necesitamos postular la existencia de una narrativa altamente estructurada que resulta incompatible con la percepción subjetiva de algunas personas. De este modo el PNC se encuentra con la imposibilidad de proponer una única entidad narrativa que posea al mismo tiempo las propiedades necesarias para poder dar cuenta de nuestras dimensiones fenomenológica y social.

2.3.2. *El Yo fenoménico*

El otro tipo de crítica que se ha lanzado para señalar la imposibilidad del PNC de atender a la dimensión subjetiva de nuestra existencia se refiere a la dificultad que tiene para proponer un sujeto de experiencias en sentido fenomenológico. Y es que si aceptamos que el Yo está constituido completamente por una narración, como hacen la mayoría de los autores, la cuestión de quién es el sujeto al que se refiere dicha narración y que experimenta por tanto de un modo consciente los hechos referidos en ella es poco clara, pues no parece posible que la propia narración pueda sentir dolor (Menary, 2008, p. 72). Cuando hablamos de experiencias conscientes está siempre implícita la idea de un sujeto de experiencias, pues no puede existir una cosa sin la otra y dicho sujeto debe ser una entidad concreta, no una historia. Siendo así, el movimiento lógico es considerar que el sujeto de dichas experiencias no es la narración en sí, sino el protagonista de la misma, pero ¿quién es el protagonista? Algunos autores como Dennett consideran que buscar un referente real para dicho protagonista es un error categorial, pues el Yo no puede ser identificado con ninguna parte de nuestro cuerpo. Desde su perspectiva el Yo no es más que una abstracción teórica, o en sus propios términos, un centro de gravedad narrativa. Es una entidad que postulamos en el centro de todas nuestras afirmaciones sobre nosotros mismos al intentar auto-interpretarnos, algo que simplifica nuestra comprensión de nuestro propio comportamiento pero que en ningún caso debe confundirse con una entidad real. Al igual que el centro de gravedad de un objeto, el centro de gravedad narrativa de una persona es una ficción teórica de gran utilidad, pero en ningún caso una entidad concreta (Dennett, 1988, pp. 103-105). Pero si adoptamos este tipo de razonamiento entonces debemos aceptar que nuestras experiencias conscientes son también ficticias, pues éstas

dependen del sujeto que las experimenta. Y aunque seguramente esto no sería un problema para Dennett, haría que las propuestas narrativas fallasen en su propósito de proponer una solución al primero de los problemas detectados en el planteamiento parfitiano de la identidad personal. Esto es, dejarían de tener en cuenta la dimensión subjetiva de nuestra existencia.²⁶

Ante esta situación la única alternativa viable es aceptar que debe existir una entidad de carácter fenoménico que es en realidad el sujeto de nuestras historias y que existe por tanto de forma previa a las mismas.²⁷ De hecho cuando experimentamos algo en primera persona, por ejemplo un dolor, no necesitamos interpretarlo de ningún modo ni situarlo en una narración para saber a quién le pertenece. Sabemos que nos pertenece a nosotros sin necesidad de ningún tipo de reflexión añadida (Zahavi, 2007, p. 188). Aunque posteriormente debamos construir una historia para poder

26 Por otro lado, si intentamos desarrollar esta postura y defender que el Yo es una ficción, acabamos teniendo que defender que las ficciones tienen algún tipo de existencia real, como se puede ver en las críticas recibidas por la teoría de Dennett (*vid.* McCarthy, 2007, pp. 73-105). El problema de Dennett parte de la confusa consideración del estatus de los elementos de ficción que maneja. Así, al mismo tiempo que nos asegura que los personajes de una novela no presentan problema alguno dado que en realidad no existen (Dennett, 1991a, p. 92) y que los Yoes sólo son otra ficción, para poder explicar su importancia en nuestra vida real también nos dice que se trata de dos tipos de ficciones diferentes (Dennett, 1991a, pp. 422-423). De este modo da a entender que hay grados entre los elementos que no existen, de forma que unos son más inexistentes que otros, lo cual parece incoherente no sólo dentro de su marco materialista, sino en cualquier ontología (Carr, 1998, pp. 339-340). Por su parte Velleman, rechazando la bizarra tesis de que existen diferentes grados de inexistencia, intenta corregir la teoría de Dennett defendiendo que al vivir acorde con la ficción de la existencia de los Yoes estos se convierten en realidad, lo que supuestamente explicaría su importancia y su papel causal en nuestras vidas. Sin embargo no parece que la solución de Velleman sea más esclarecedora que la de Dennett, pues igual que no sabemos en qué podrían consistir los grados de inexistencia tampoco queda claro como una entidad puede ser existente e inexistente al mismo tiempo (Velleman, 2005, p. 221). En consecuencia nos vemos forzados a admitir que si defendemos que el Yo es una ficción tenemos que concederle algún tipo de existencia real a las ficciones, lo que implica abandonar el materialismo y admitir que en el mundo hay entidades que no son físicas.

27 Señala David Carr criticando a Dennett que una historia no es tal hasta que alguien la interpreta, por lo que un Yo narrativo ficticio dependería en última instancia de un Yo no-narrativo real que interpretase las historias como tales y les diese sentido (Carr, 1998, p. 342).

articular dicha experiencia y comprenderla, existe un sentido previo en el que es experimentada por nosotros. Este sujeto de experiencias con un carácter pre-narrativo y pre-lingüístico ya era reconocido en el ámbito de la fenomenología clásica por autores como Husserl, Sartre o Merleau-Ponty. Pero no sólo por ellos, sino que de forma más reciente el neurólogo Antonio Damasio también ha distinguido entre un Yo central, que se relaciona con un primer nivel de conciencia de nuestro entorno que estaría presente no sólo en los seres humanos sino también en otros animales, y un Yo autobiográfico, que depende del tipo de auto-conciencia que alcanza su máximo grado de desarrollo en nosotros gracias al uso del lenguaje (Damasio, 2010, pp. 277-279).²⁸ El caso de los enfermos de Alzheimer puede servir como ejemplo para entender esta distinción, pues aunque en las fases más avanzadas de la enfermedad los sujetos carecen de la capacidad de adoptar una postura reflexiva sobre sí mismos y articular una narración de su vida, nada parece indicar que sean incapaces de sentir dolor y experimentarlo como propio y no como un dolor impersonal (Zahavi, 2007, pp. 192-193).

Para simplificar nuestra terminología Dan Zahavi sugirió que podríamos denominar «Yo» a esta entidad mínima relacionada con el nivel más básico de la conciencia de la que estamos hablando y «persona» a la entidad resultante de la construcción narrativa (Zahavi, 2007, p. 193). De este modo la primera entidad podría dar cuenta de la dimensión subjetiva de nuestra existencia, mientras que la segunda haría lo propio con nuestra dimensión social. Como veremos a continuación, ésta solución coincide en líneas generales con la que Schechtman dio como respuesta a las críticas de Strawson, de tal modo que la distinción entre Yoes y personas ha sido ampliamente aceptada entre los teóricos del Yo narrativo dando lugar al NPN.

2.4. El Nuevo Planteamiento Narrativo: problemas sin resolver

Como decíamos, Schechtman ofreció una respuesta a Strawson mucho más prometedora que las de Rudd y Stokes que supondría hacer compatibles las críticas a las dos tesis de la Narratividad con su propia visión sobre la identidad narrativa. Señala Schechtman que la tensión subyacente entre los aspectos fenoménicos y sociales de nuestra existencia

28 Damasio también habla de la existencia de un proto Yo que estaría relacionado con la información que recibe el cerebro sobre el estado del propio organismo, pero para el caso podemos simplificar su posición y quedarnos con el Yo central y el autobiográfico.

que se da en el PNC podría ser resuelta si, como hace Zahavi, diferenciásemos entre Yoes –en tanto sujetos de experiencias en un sentido fenomenológico– y personas –en tanto entidades existentes en un contexto social–. Al hacerlo podríamos aceptar que un Yo no necesita tener una duración temporal extendida ni una comprensión narrativa de sí mismo más que en un sentido trivial, tal y como Strawson había defendido, lo que no nos impediría asegurar al mismo tiempo que una persona necesita de una estructura narrativa que se extienda a lo largo de toda la vida del ser humano que también es para poder dar cuenta de los aspectos morales (Schechtman, 2007, pp. 168-169).²⁹ De este modo las teorías narrativas todavía podrían encontrar una solución al problema de la identidad personal.

La coincidencia entre la propuesta de Zahavi y la solución a las críticas de Strawson que ofreció Schechtman sugieren que una alternativa de este tipo podría ser el camino que debería seguir la investigación sobre el problema de la identidad narrativa. Aún sería necesario un cierto trabajo para poder unificar las propuestas de Schechtman y Zahavi, ya que la primera atribuye un grado de narratividad al Yo que está ausente en la formulación del segundo.³⁰ Por otro lado no queda claro cómo las dos tesis de la Narratividad, que tienen un profundo carácter psicológico, pueden aplicarse a entidades de carácter social como son las personas, pero sin embargo ninguno de estos dos problemas parece un obstáculo insalvable. De este modo el NPN se ofrece como un nuevo marco, todavía por desarrollar, en el que mediante la diferenciación entre Yoes y personas sería posible solucionar el problema de la identidad personal.

Pero el principal inconveniente, en mi opinión, al que se debe enfrentar el NPN es el tipo de consideración que deberían tener el Yo y la persona, ya sea en relación uno a la otra o ambos en relación al ser humano que también somos. Especialmente en lo que respecta a nuestra supervivencia y a las otras tres características de la identidad personal

29 Ya Dennett muchos años antes había señalado que existían al menos dos sentidos del concepto de «persona», uno metafísico y uno moral, y que ambos no eran necesariamente idénticos. Sin embargo mantenía que seguramente ser persona en el sentido metafísico (esto es, ser un Yo) era una condición necesaria para ser persona en el sentido moral, y que ambos sentidos podían ser entendidos mejor, más que como dos conceptos distintos, como dos bases de apoyo de un mismo continuo, por lo que no se molesta en diferenciarlos (Dennett, 1976, pp. 7-8, 36).

30 La propia Schechtman sugiere una posible solución, argumentando que Zahavi y ella podrían estar refiriéndose a dos tipos de conciencia distintos (Schechtman, 2011, pp. 410-411). En cualquier caso volveré sobre esto más adelante.

señaladas por Schechtman, pues las opiniones sobre este punto de los distintos autores varían enormemente. Así por ejemplo parece deducirse de lo que hemos dicho de las posturas de Zahavi y Schechtman que la existencia de personas implica la existencia de Yoes, pero no al revés, lo que podría llevarnos a suponer que nuestra existencia como Yoes debería importarnos más que nuestra existencia como personas, aunque Zahavi niega explícitamente esta conclusión (Zahavi, 2007, pp. 193-194). Por otro lado Schechtman señala que se debería tomar como un deber moral el intentar que nuestro Yo tenga una duración temporal tan extensa como la de la persona, y que ésta a su vez prolongue su existencia hasta abarcar la totalidad de la vida del ser humano que también somos, de modo que las tres entidades sean coextensivas, pues teniendo en cuenta que lo que nos debe importar cuando nos preocupamos por las cuatro características de la identidad personal son las personas, si la duración de las tres entidades no coincide se pueden producir problemas tanto emocionales como éticos (Schechtman, 2007, pp. 175-178). Por otro lado Richard Menary señala que la idea de un Yo es incoherente a no ser que se considere como producto de una conciencia in-corporada, por lo que en el fondo tanto el Yo como la persona dependen del ser humano que somos en última instancia y cuya supervivencia es la que nos debe importar (Menary, 2008, pp. 73-79).

Lo que estos desacuerdos revelan, en mi opinión, es la necesidad de clarificar el marco ontológico que subyace al NPN, así como los significados de «ser humano», «Yo» y «persona» que se derivan de él. Sin hacerlo no podemos decir qué relaciones se dan entre tres entidades de distinta naturaleza pero que comparten el mismo espacio físico y cuál o cuáles nos deben importar cuando nos preocupamos por las cuatro características. Para poder responder a esto primero debemos ser capaces de decir en qué sentido –si es que en alguno– la existencia de unas entidades depende de otras; si esta dependencia implica que unas entidades tienen un estatus ontológico más fundamental; o si la aparición de entidades en un nivel ontológico más elevado implica la existencia de propiedades emergentes que no se pueden justificar desde un nivel inferior y que podrían apoyar la tesis de que lo que nos debe importar en nuestra supervivencia no es la supervivencia de la que seamos en un nivel más fundamental, sino la que somos en un nivel más elevado. En definitiva, deberemos abordar el problema de cómo se relaciona el aspecto físico del mundo con la conciencia y con las entidades sociales. En cualquier caso no es mi intención resolver un problema tan complejo en este capítulo, pues le dedicaré más atención en los dos siguientes. Sin embargo sí me gustaría señalar que, para dar una explicación satisfactoria de la identidad

personal, creo que el NPN deberá abandonar no sólo el reduccionismo, como ya hizo el PNC, sino también el materialismo.

2.5. Conclusiones

El PNC se propuso como un enfoque alternativo al PO que habían desarrollado Parfit, Williams, Shoemaker y otros durante la segunda mitad del siglo XX sobre el problema de la identidad personal a través del tiempo. Y dicho enfoque partía con la intención, mediante las dos tesis de la Narratividad, de tener en cuenta dos aspectos de nuestra existencia que habían sido ignorados hasta el momento por la postura reduccionista que dominaba el PO. Por un lado la perspectiva radicalmente subjetiva desde la que consideramos el mundo. Por otro la imposibilidad de comprender nuestra vida si no se atiende al hecho de que vivimos necesariamente en sociedad. Para ello propusieron que nuestra identidad podría consistir en nuestro relato subjetivo sobre nuestra propia vida (tesis psicológica de la Narratividad), relato que debía darse en un contexto social que le impondría una serie de requisitos (tesis ética de la Narratividad).³¹ De este modo estaríamos teniendo en cuenta el modo en que nuestra conciencia percibe el mundo, al mismo tiempo que al no poder desviarse nuestro relato del marco social en el que nos desarrollamos también estaríamos atendiendo a esta segunda faceta de nuestra existencia.

Sin embargo, tras más de dos décadas de reinado narrativista, recientemente han aparecido múltiples críticas a este enfoque que en el fondo demuestran la incapacidad del planteamiento propuesto por MacIntyre, Taylor, Ricoeur y otros para resolver las dos carencias que ellos mismos habían detectado en el reduccionismo del PO. Y es que si intentamos que nuestra narración se adapte lo más posible a nuestra subjetividad acabamos teniendo que renunciar a toda perspectiva narrativa, al menos en cualquier sentido significativo del término que nos permita tener en cuenta al mismo tiempo nuestra vida social. Si por el contrario intentamos que nuestra narración tenga las características apropiadas para dar cuenta de la dimensión social de la identidad personal, necesitamos postular la existencia de una narrativa altamente estructurada que resulta incompatible con nuestra percepción subjetiva. De este modo el PNC se encuentra con la imposibilidad de proponer una única entidad narrativa

31 Como hemos visto, entre los requisitos que impone la tesis ética de la Narratividad para algunos autores como MacIntyre se encuentra la necesidad de tener en cuenta los relatos de los otros. En cualquier caso, como ya he dicho, volveré sobre esto en el quinto capítulo.

que posea al mismo tiempo las propiedades necesarias para poder dar cuenta de nuestras dimensiones fenomenológica y social.

La solución a este problema aún está por desarrollarse, aunque entre las respuestas ha surgido con fuerza la opción de establecer una diferenciación entre lo que somos como Yo y lo que somos como persona, de modo que cada una de estas dos entidades pueda dar cuenta de uno de estos dos aspectos de nuestra existencia. Esta posible solución es la base de lo que denomino el NPN. Personalmente me inclino por considerar que la investigación sobre el problema de la identidad personal debe seguir por esta línea, y a ello dedicaré las próximas páginas. Sin embargo como hemos podido ver no queda claro qué marco ontológico podría dar un soporte adecuado al NPN, y sin él resulta difícil saber qué entidad o entidades nos debería importar cuando nos preocupamos por nuestra supervivencia o por la responsabilidad moral. En consecuencia, antes de desarrollar este planteamiento y definir de un modo más exhaustivo en qué consiste la diferenciación entre Yoes y personas será necesario clarificar nuestros presupuestos ontológicos. Para ello primero analizaré por qué considero que el NPN no sólo es incompatible con el reduccionismo, sino también con el materialismo, para a continuación proponer un marco alternativo.

CAPÍTULO 3

LOS PROBLEMAS DEL MATERIALISMO

Como hemos venido señalando desde la introducción, el debate sobre la identidad personal ha estado unido desde sus orígenes a la pregunta acerca de qué es lo que nos debe importar cuando nos preocupamos por nuestra supervivencia y por la responsabilidad moral. Además, desde que Locke lo estableciese así, existía el acuerdo tácito de que el término «persona» debía utilizarse para referirse a la entidad que estuviese vinculada con estos dos aspectos. Según los animalistas esta entidad era el ser humano, mientras que para los neo-lockeanos esta entidad era el Yo. Sin embargo, como vimos al final del capítulo anterior, en el NPN se decide utilizar el término «persona» de forma novedosa para referirse a una entidad de carácter social que debe distinguirse tanto del ser humano como del Yo, lo que ha provocado que el debate sobre qué entidad es la que nos debe importar cuando nos preocupamos por nuestra supervivencia o por la responsabilidad moral se complique aún más. En este sentido, a pesar de que los partidarios del PNC podían considerarse neo-lockeanos, pues defendían que lo que nos debía importar en nuestra supervivencia era el Yo, en la evolución del enfoque narrativo que supone el NPN no hay ningún acuerdo al respecto.

Como resulta evidente, para poder tomar una decisión sobre qué entidad es la que nos debe importar cuando nos preocupamos por la supervivencia y la responsabilidad moral primero hemos de tener unas definiciones más concretas de los conceptos de «ser humano», «Yo» y «persona». Sin embargo para ello necesitamos aclarar también el marco ontológico en el que se va a desarrollar el NPN, lo que parece conducimos

a un círculo vicioso, pues ¿cómo podríamos saber cuál es el marco ontológico apropiado para el NPN si no sabemos qué es lo que se quiere sostener en él? Pero a pesar de que no tenemos unas definiciones que puedan considerarse definitivas de las tres entidades que somos, sí contamos con algunas aproximaciones e intuiciones al respecto desde las que partir para elaborarlas.

Siendo así, tomando como punto de arranque lo que ya sabemos sobre el NPN, en el primer apartado de este capítulo presentaré unas definiciones preliminares de lo que creo que deberíamos entender por «ser humano», «Yo» y «persona». Armado con el esbozo de estos conceptos, pondré a prueba nuestras intuiciones acerca de qué es lo que nos debe importar cuando nos preocupamos por la supervivencia o por la responsabilidad moral, lo que me permitirá sostener tres tesis acerca de lo que somos:

1. Que las condiciones de supervivencia de las tres entidades que se distinguen en el NPN son independientes entre sí, por lo que la muerte de cualquiera de ellas no implica nada acerca de la supervivencia de las otras dos.
2. Que cuando hablamos de supervivencia, nos importan en la misma medida las tres entidades.
3. Que cuando hablamos de responsabilidad moral sólo nos importan las personas.¹

1 En ocasiones la discusión sobre qué entidad es la que nos debe importar cuando nos preocupamos por nuestra supervivencia y por cuestiones de responsabilidad moral adopta la forma de la pregunta por la entidad que somos *fundamentalmente* (vid. Baker, 2000, pp. 3-12). Y en versiones previas de este trabajo yo mismo adopté esa terminología. Sin embargo ahora considero que se presta a confusiones evitables, pues parece implicar que existe una unidad primordial de aquello que somos, lo cual choca con las tesis que pretendo defender. Presupongamos, como sin duda es posible hacer, que la entidad que *soy* fundamentalmente es aquella sin la cual *yo* no existiría en absoluto, y que por tanto su existencia es una condición necesaria y suficiente para *mi* supervivencia. Si esto es así, parece no ser posible la defensa de que somos fundamentalmente tres entidades cuyas condiciones de supervivencia son independientes, pues si la muerte de una de ellas no *me* impidiese sobrevivir como las otras dos entonces sus respectivas existencias no serían condiciones necesarias para *mi* supervivencia. Como digo se trata de una confusión evitable, pues la contradicción sólo es aparente. El lenguaje nos limita al hacernos hablar desde la primera persona del singular, dando la sensación de que existe un único referente de «yo». Sin embargo lo que en el fondo pretendo demostrar es que no existe tal unidad, sino que cuando me pregunto por *mi* supervivencia en realidad me estoy preguntando por la supervivencia de tres

Una vez haya defendido que nuestras intuiciones acerca de lo que somos y lo que nos debe importar cuando nos preocupamos por la supervivencia o por la responsabilidad moral quedan reflejadas en estas tesis, en los dos apartados restantes mostraré que el materialismo no es el marco apropiado donde desarrollar el NPN. En el primero de ellos haré explícitos los problemas que tienen los materialistas para explicar las propiedades fenoménicas. Dado que definiré el concepto de «Yo» atendiendo precisamente a su consideración como sujeto de experiencias, las enormes dificultades que tiene el materialismo para explicar este aspecto de nuestra existencia creo que constituyen un motivo suficiente para su abandono. No obstante el grueso de la argumentación recaerá en el último apartado, donde trataré de demostrar que resulta imposible justificar en un marco materialista que debemos preocuparnos por la supervivencia de las personas en el mismo grado en el que nos preocupamos por la supervivencia de los seres humanos y los Yoés.

En consecuencia, si lo que pretendo defender en este capítulo es cierto, sólo nos restarán dos opciones: o abandonar nuestras intuiciones acerca de lo que somos y lo que nos debe importar en nuestra supervivencia, o renunciar al materialismo. Como he venido anunciando desde el principio, me decantaré por esta segunda posibilidad, por lo que en el próximo capítulo presentaré un marco alternativo que sí nos permitirá justificar nuestras intuiciones y a partir del cual podrán desarrollarse con más precisión los conceptos de «ser humano», «Yo» y «persona». En cualquier caso un aspecto que merece nuestra atención antes de seguir adelante es que hasta el momento he venido hablando de «materialismo» sin entrar en mayores precisiones. Sin embargo si, como voy a tratar de defender, es necesario abandonar esta postura para resolver el problema de la identidad personal en el marco del NPN entonces primero debemos precisar un poco más qué estoy entendiendo con ello.

Aunque hay muchas versiones del materialismo, a grandes rasgos podemos decir que el materialismo estándar es el marco que se deriva de aceptar por un lado la teoría atómica de la materia y por otro la teoría de la

entidades distintas, por «nuestra» supervivencia. Siendo así, la supervivencia de cada una de estas entidades es una condición suficiente y necesaria para la supervivencia de alguna parte de ese «nosotros» que llamamos «yo». En cualquier caso, aunque la confusión podría aclararse, he preferido abandonar esa terminología y hablar únicamente de qué entidad es la que nos preocupa cuando hablamos de supervivencia o de responsabilidad moral. (Debo agradecer sus comentarios a quienes me escucharon hablar sobre este tema en sendos seminarios en la Australian National University y en la Universidad Nacional Autónoma de México, especialmente a Aliosha Celeste Barranco López).

evolución de las especies: todo lo que existe está compuesto por partículas físicas fundamentales, cuya combinación produce todos los fenómenos del mundo real, incluidas tanto la vida como la experiencia consciente (Searle, 1992, pp. 97-103). Esta afirmación además se complementa con la creencia de que la ciencia física actual es en líneas generales verdadera –según una teoría de la verdad como correspondencia con la realidad–, por lo que si todo lo que existe es físico entonces puede ser reducido a una explicación en términos de la ciencia física actual o una versión no-revolucionaria de la misma (Strawson, 1994, p. 19, 2006, p. 4). Si algo no puede explicarse en términos físicos entonces no existe, no es real.

Esta caracterización del materialismo, ampliamente aceptada entre los filósofos analíticos de la mente durante las últimas décadas, se enfrenta a serios problemas a la hora de explicar las experiencias conscientes, pues no parece claro en qué sentido la conciencia puede ser física cuando desde Descartes la filosofía moderna se ha construido sobre la oposición entre lo físico y lo mental. De este modo todos los intentos de explicar la conciencia dentro de un marco materialista –conductismo, teoría de la identidad, funcionalismo (*vid.* Moya, 2004)...– se han encontrado con una barrera insalvable. Y es que por muy desarrollada que pueda estar una teoría de la conciencia en términos neuronales o funcionales, por mucho que su capacidad predictiva sobre los fenómenos conscientes sea muy elevada –incluso podemos conceder que aunque sea completa–, siempre parece existir un vacío imposible de llenar, una brecha entre lo físico y lo mental que nos impide aceptar una identidad –sea del tipo que sea– entre, por ejemplo, la estimulación de las fibras nerviosas de tipo C y el dolor que experimentamos de forma subjetiva. Parece existir un *vacío explicativo* entre la dimensión subjetiva de la experiencia y la causa física propuesta como explicación por cualquiera de las teorías materialistas (*vid.* Levine, 1983). Aun cuando la estimulación de las fibras C y las sensaciones conscientes de dolor aparezcan invariablemente unidas, parece que aún falta una cosa por explicar: ¿por qué la estimulación de las fibras C debe sentirse como dolorosa cuando existen los mismos motivos para que se sienta como placentera, o para que no se sienta en absoluto?

La solución a este problema en términos estrictamente materialistas parece abocarnos al eliminativismo, esto es, a la creencia de que los qualia –las cualidades fenoménicas– no existen, pues no son reducibles a la física (*vid.* Churchland, 1984).² Esto ha llevado en los últimos años a la

2 A este respecto podemos recordar que aunque el reduccionismo parfitiano no es equivalente al reduccionismo de lo mental a lo físico que proponen las posturas eliminativas, existe una cierta afinidad, como señalé en el capítulo anterior (*vid.* Shoemaker, 1985, p. 139).

aparición de teorías que, sin abandonar el materialismo, tratan de hacer compatibles lo mental y lo físico mediante una reformulación de nuestros presupuestos ontológicos, ya sea proponiendo alguna forma de emergentismo o pansiquismo. Sin embargo, como trataré de mostrar, estas posturas materialistas tampoco pueden considerarse un marco apropiado en el que desarrollar el NPN, pues por un lado no tienen éxito en su propósito de explicar las propiedades fenoménicas, y por otro tampoco pueden ofrecer una justificación que podamos considerar aceptable de nuestras intuiciones acerca de qué es lo que nos debe importar cuando nos preocupamos por nuestra supervivencia.

3.1. Lo que nos importa

Como acabo de señalar, la tesis principal de este capítulo es mostrar que el problema de la identidad personal no puede solucionarse en un marco materialista. Pero para hacerlo primero he de defender otras tres tesis: (1) que las condiciones de supervivencia de las tres entidades que se distinguen en el NPN son independientes entre sí, por lo que la muerte de cualquiera de ellas no implica nada acerca de la supervivencia de las otras dos; (2) que cuando hablamos de supervivencia, nos importan en la misma medida las tres entidades que somos; y (3) que cuando hablamos de responsabilidad moral sólo nos importan las personas. La idea es que estas tres tesis son las que mejor dan cuenta de nuestras intuiciones sobre el tema y que ninguna variante del materialismo puede justificarlas.

Sostener las tres tesis que he enunciado es complicado cuando aún ni siquiera contamos con unas definiciones medianamente claras de lo que es un ser humano, lo que es un Yo y lo que es una persona. Sin embargo, aunque no tenemos unas descripciones definitivas de lo que somos, sí contamos con algunas aproximaciones e intuiciones al respecto. En este sentido para defender las tres tesis presentaré unas definiciones preliminares de lo que creo que deberíamos entender por «ser humano», «Yo» y «persona» a partir de lo que expuse en el capítulo anterior para posteriormente ver que lo que pensamos intuitivamente acerca de cuál de ellas debería importarnos cuando nos preocupamos por nuestra supervivencia o por cuestiones de responsabilidad moral queda recogido en ellas.

3.1.1. Aproximación conceptual

El concepto de «ser humano» es el menos problemático de los tres que nos ocupan y no me detendré en él, pues en líneas generales podemos

aceptar la definición que la biología da de él como un animal con ciertas características físicas y que puede desarrollar una serie de capacidades. En cuanto a las definiciones de «Yo» y «persona», es importante recordar primero que, como vimos en el anterior capítulo, la distinción entre ambos conceptos surge a raíz de dos discusiones diferentes y que por tanto no todos los autores coinciden en la forma que se debe dar a la misma. Por un lado filósofos como David Carr, Dan Zahavi o Richard Menary señalan la dificultad que tiene el PNC, si quiere defender que el Yo está constituido íntegramente por una narración, para postular un sujeto de experiencias en sentido fenomenológico. Si la narración habla de un dolor de muelas, ¿quién es el sujeto que siente ese dolor? ¿La propia narración? Ante esta dificultad defienden que debe distinguirse entre un Yo fenomenológico, mínimo, que sea el sujeto de las experiencias y que sea previo a todas las narraciones; y un Yo social o persona que sea fruto de la construcción narrativa posterior que este Yo fenomenológico hace en algunos casos para comprender su vida extendida en el tiempo (por ejemplo los enfermos de Alzheimer tendrían un Yo fenomenológico, pero no serían personas ya que serían incapaces de construir una narración de su vida). Por otra parte, Marya Schechtman propuso distinguir entre Yoes y personas para responder a las críticas realizadas por Galen Strawson contra el PNC, en las que señalaba que algunos sujetos, a pesar de tener la capacidad de comprender adecuadamente sus vidas, no las unifican mediante una narración ni de ningún otro modo (Strawson, 2004). Ante esta situación Schechtman sugirió la posibilidad de diferenciar entre un Yo trivialmente narrativo y más vinculado a nuestra experiencia fenomenológica, y una persona completamente narrativa de existencia más amplia que atendiese al hecho de que vivimos necesariamente en sociedad. Aunque así resumidas las dos propuestas parecen similares, en realidad lo que Zahavi llama «persona» es más parecido a lo que Schechtman llama «Yo» que a lo que llama «persona», de modo que existe aquí un desacuerdo sobre el que es necesario hacer aclaraciones. Mi postura es más cercana a la de Schechtman, por lo que será en ella en la que me base. En cualquier caso de momento no explicaré hasta qué punto es importante la diferencia entre las posturas de Zahavi y Schechtman, ni argumentaré tampoco el porqué de mi decisión.³

Teniendo en mente esta aclaración previa, podemos decir que el Yo hereda la mayor parte de las características que se le habían asignado desde Locke. De este modo podemos considerar preliminarmente que el Yo es una entidad mental caracterizada por ser el sujeto de las experiencias

3 Para ello véase el capítulo 5, en el que desarrollo las definiciones preliminares que ofrezco aquí.

conscientes, ser un agente racional y tener autoconciencia.⁴ Al mismo tiempo diremos que un Yo es una entidad distinta del cerebro que habita, aunque no posee independencia con respecto al mismo, pues debe estar encarnado en el cerebro concreto que lo genera y sustenta. Por último podemos ver que dado que los Yoes son un ente diferenciado de los seres humanos por cuyos cerebros son generados, su existencia no tiene porqué coincidir con la de los mismos. La vida de un Yo puede ser mucho más reducida que una vida humana normal –según Strawson un Yo Episódico no tendría porqué durar más de lo que dura una experiencia fenoménica aislada– y también más amplia –un Yo podría sobrevivir teniendo como base un cerebro dentro de una cubeta–. En cualquier caso diré algo más sobre esto en las próximas páginas.

Finalmente las personas son la entidad que menos definida se encuentra hasta la fecha en el NPN, pues apenas contamos con algunas ideas de Schechtman al respecto. Sabemos que es una entidad que se construye narrativamente, y que su existencia es por norma más amplia que la de los Yoes, tendiendo a coincidir con la de los seres humanos, aunque no sea necesario. Además sabemos que aparece en un contexto social, y que el principal motivo para postular su existencia es precisamente dar cabida a la dimensión social de nuestra vida. Ante esta situación, me gustaría fijarme para desarrollar un poco más el concepto en las ideas de MacIntyre sobre el Yo, pues este autor abrió una puerta sumamente interesante para introducir la dimensión social de las personas en el debate: como dije en el capítulo anterior, para MacIntyre nuestra identidad no está constituida únicamente por el relato que contamos sobre nuestras vidas, sino también por los relatos que sobre nosotros hacen los que nos rodean, de modo que no somos más que meros coautores de nuestra historia (MacIntyre, 1981, p. 263). Aunque esta idea necesita todavía más desarrollo, de momento podemos basarnos en lo anterior para decir que las personas son entidades sociales construidas narrativamente a partir de creencias colectivas acerca de nuestra existencia, creencias que son fruto de la negociación entre nuestro relato de vida y los relatos que cuentan los demás.

Sin duda estas definiciones son todavía demasiado provisionales y débiles como para afrontar de pleno el problema de la identidad personal. Soy consciente además de que una tesis audaz como la que pretendo

4 Esta definición tiene mucho que ver con la que Locke da de «persona» en el *Ensayo sobre el entendimiento humano*: «Y es [la persona], me parece, un ser pensante inteligente dotado de razón y de reflexión, y que puede considerarse a sí mismo como el mismo, como una misma cosa pensante en diferentes tiempos y lugares» (Locke, 1690, p. 318).

defender aquí requiere de una demostración más rigurosa que una tesis más complaciente. Pero como decía más arriba por el momento sólo me interesa señalar estos rasgos preliminares y ponerlos a prueba para sacar a la luz algunas de nuestras intuiciones sobre lo que somos en tanto seres humanos, Yoes y personas. En concreto lo que me gustaría demostrar a continuación es que podemos identificarnos con las tres entidades que somos en la misma medida, que podemos sobrevivir como cualquiera de ellas de forma independiente a lo que nos ocurra como las otras dos y que cuando nos preocupamos por la responsabilidad moral únicamente nos importan las personas. Y para ello me gustaría recurrir a un experimento mental que mezcla dos de los experimentos más famosos del debate que se dio sobre la identidad personal en el PO: el caso del trasplante de cerebro y el caso del teletransporte.

3.1.2. El jefe mafioso

Podemos imaginar a un peligroso jefe de la mafia de New Jersey, llamémosle Tony Soprano, que es detenido y acusado de múltiples asesinatos. Las pruebas son sólidas, así que nadie duda de que será condenado a muerte. El caso es tan sencillo que el fiscal ni siquiera necesitará preparar el juicio para ganarlo. Consciente de este hecho, el abogado defensor va a visitar a Tony y le confiesa que por mucho que pueda retrasar la sentencia es imposible que le libre de una condena a muerte, pero que sin embargo existe una manera de multiplicar por tres sus posibilidades de supervivencia. Se trata de un procedimiento experimental mediante el que un científico primero creará una réplica exacta de su cuerpo que por tanto será física y psicológicamente idéntica a él en términos cualitativos, al igual que sucede en el caso del teletransporte. Posteriormente el científico tomará su cuerpo original y extraerá de él la parte superior del cerebro para conectarla a un dispositivo robótico que, cuando despierte de la operación, le permitirá ver, oír y comunicarse, aunque no moverse. Finalmente, tomará el resto de su cuerpo y lo conectará a una serie de máquinas que lo mantendrán con vida, aunque al carecer del cerebro superior nunca podrá recobrar la conciencia. De este modo, tras la operación existirán tres entidades que podrían ser consideradas Tony Soprano. El abogado le dice que en el juicio sólo pueden condenar a una de esas entidades, pero que sea cual sea la que sufra la condena a muerte, Tony tendrá motivos para alegrarse de su futuro.

Según vimos en el capítulo dedicado al PO, en su mayoría los animalistas defenderían que lo que debería importarle al jefe mafioso es la

supervivencia de su cuerpo original, que ha quedado en coma. Los argumentos presentados por Williams acerca de la preocupación que podemos sentir hacia el futuro de nuestro cuerpo independientemente de la continuidad psicológica apoyarían esta posición. Por su parte los neo-lockeanos –tanto los partidarios del criterio psicológico del PO como los narrativistas del PNC– defenderían que el jefe mafioso debería preocuparse por el futuro de su cerebro, que ahora está conectado a un robot incapaz de moverse aunque mentalmente activo. La existencia de continuidad mental y que ésta sea producida por la existencia continua de un mismo cerebro serían los puntos fuertes de su argumentación. Por último la réplica no tendría derecho a reclamar la identidad del jefe mafioso ni para los animalistas ni para los neo-lockeanos, a pesar de que como Parfít señaló tendría todo lo que realmente debería importarle en su supervivencia (*vid.* Parfít, 1984). En este sentido podemos ver que existen argumentos que podrían apoyar la afirmación del abogado de que el mafioso tendrá motivos para alegrarse sea cual se la entidad que sufra la condena a muerte. El único problema para defender esto podría ser la creencia de que existe una unidad primordial en aquello que somos, lo que podría inducirnos a pensar que sólo podemos preocuparnos por el futuro una de las tres entidades, o que como mínimo la supervivencia de una de las tres nos debe importar más que la de las otras dos. Sin embargo, como señalé anteriormente en la primera nota al pie del capítulo, esta creencia parece estar fundada en una confusión de orden lingüístico que nos hace pensar que cuando utilizamos el pronombre «yo» nos referimos siempre a una misma y única entidad. Pero si renunciamos a la idea de que existe dicha unidad previa, no existe ningún motivo por el que Tony Soprano no pueda identificarse en la misma medida con las tres entidades resultantes del experimento y preocuparse igualmente por su supervivencia.

Por otro lado, si analizamos el experimento teniendo en cuenta la reciente diferenciación que se da en el NPN entre lo que somos como seres humanos, Yoes, y personas, podemos ver en primer lugar que ni el robot ni la réplica podrían aspirar a ser el mismo ser humano que era Tony Soprano antes de la operación, pero que sin embargo el cuerpo en coma sí puede ser considerado como tal. En segundo lugar, el robot contiene el cerebro del mafioso intacto, por lo que se puede considerar que posee la misma conciencia que le definía en tanto Yo, mientras que ni el cuerpo en coma –que ni siquiera posee cualidades mentales– ni la réplica –no se puede decir que tengan las mismas propiedades mentales si éstas proceden de bases físicas distintas (*vid.* Williams, 1957)– podrían aspirar a dicha consideración. Finalmente la réplica no sólo contiene todo lo que debería importarle al jefe mafioso en su supervivencia, sino que si nos ponemos

por un momento en la piel del juez podemos ver que no tendría sentido meter en la cárcel al cuerpo en coma ni al robot incapaz de moverse, pero que sin embargo sí lo tendría meter a la réplica, pues ésta podría asumir el mismo rol social que tenía Tony Soprano antes de la operación y seguir cometiendo el mismo tipo de crímenes. En este sentido desde el NPN la réplica podría ser considerada la misma persona que el jefe mafioso y ser por tanto la responsable legal de sus acciones ante la sociedad. Podemos entonces decir que la operación conseguirá separar las tres entidades que el jefe mafioso era anteriormente, de modo que sobrevivirá en tanto ser humano gracias a su cuerpo en coma, en tanto Yo gracias a su cerebro conectado a un robot y en tanto persona gracias a su réplica. Lo que antes eran tres entidades que coincidían en un único soporte físico tras la operación seguirán existiendo pero separadas en tres cuerpos diferentes.

Tras exponerle el caso, Tony puede darse cuenta de que, si se somete a la operación, suceda lo que suceda en el juicio de algún modo morirá. Pero al mismo tiempo también sobrevivirá. Sin duda preferiría seguir viviendo de una sola pieza, pero si tiene que morir es mejor hacerlo sólo en parte. Si condenan a su cuerpo sabe que morirá como ser humano, pero sin embargo puede alegrarse al saber que en tanto Yo podrá seguir disfrutando de su afición a las películas de *gangsters*, mientras que en tanto persona podrá seguir dirigiendo su familia del mismo modo a cómo lo ha hecho hasta ese momento. Si por el contrario condenan al robot, morirá como Yo pero seguirá dirigiendo a su familia y como ser humano podrá seguir teniendo hijos o esperando un futuro milagro médico-tecnológico en forma de prótesis cerebral. Finalmente, si condenan a la réplica morirá como persona, pero seguirá existiendo como un Yo con todos los placeres mentales a su alcance y como ser humano estará biológicamente vivo. Aunque ajusticiasen a dos de las entidades en lugar de a una seguiría siendo un buen trato.

En este sentido podemos ver que el experimento mental del jefe mafioso apoya las tres tesis cuya defensa anuncié al principio del capítulo. En primer lugar, muestra que las condiciones de supervivencia de seres humanos, Yoes y personas no están relacionadas entre sí, pues independientemente de a qué entidad decidan condenar en el juicio, las otras dos sobrevivirán. Además presenta argumentos a favor del hecho de que nos pueden importar en la misma medida la supervivencia de las tres entidades que somos, pues como hemos dicho Tony tendría motivos para alegrarse fuese cual fuese el resultado del juicio. Y por último, también muestra que cuando hablamos de responsabilidad moral sólo nos importan las personas, pues es precisamente el hecho de que la réplica de Tony podría desempeñar el mismo rol social que el mafioso lo que nos induce a

pensar que el juez debería considerarla como la responsable de sus acciones.

En este momento alguien podría decir que estaría dispuesto a aceptar que en una situación así Tony Soprano sobreviviría como ser humano y como Yo –al fin y al cabo, como acabamos de decir, los partidarios de los criterios físicos y psicológicos del PO defendían algo así y ambos daban buenas razones–, pero que su supervivencia como persona no estaría tan clara. Y que en caso de darse, no es evidente que el Tony Soprano que existe antes del experimento pudiera identificarse con su réplica futura en el mismo grado en el que se identifica con lo que es como ser humano y como Yo, hasta el punto de poder alegrarse si sólo «sobreviviese» como persona gracias a su réplica.

Ambas objeciones se responden del mismo modo. Como hemos dicho antes, las personas son entidades sociales construidas narrativamente a partir de creencias colectivas acerca de nuestra existencia. Creencias que son fruto de la negociación entre nuestro relato de vida y los relatos que cuentan los demás. En este sentido para que Tony Soprano sobreviva como persona gracias a su réplica hacen falta dos factores: el primero es que el propio Tony Soprano esté dispuesto a otorgarle su identidad como persona a su réplica (que su narración sobre sí mismo contemple a su réplica como a su persona futura); el segundo, es que la sociedad en la que sucede el caso tenga como cierta la creencia de que la réplica es la misma persona que era Tony Soprano, lo cual hemos dicho que sería oportuno dado que el juez desearía poder considerar a la réplica como la responsable de las acciones pasadas del jefe mafioso. Si se cumplen ambas condiciones, Tony se identificará con su réplica –pues así lo establece su narración–, mientras que dicha identificación será aceptada socialmente, por lo que la negociación entre la narración de Tony y las narraciones de otros tendrá como resultado que la réplica será la misma persona que era Tony Soprano antes del experimento.

Alguien podría argumentar también que una identificación de este tipo sin duda podría basarse en falsas creencias, y que del mismo modo que una sociedad podría aceptar que la réplica de Tony Soprano es la misma persona que el mafioso original, también podría darse el caso de que todos estuviesen de acuerdo en que su identidad podría ser transferida a otra persona cualquiera, o incluso a una piedra. Si esto fuese posible entonces la condición de persona no tendría ningún valor, y difícilmente podría ser significativo para Tony sobrevivir de ese modo.

Ante esta crítica debemos profundizar en algunos aspectos de la definición de «persona» que di anteriormente. En primer lugar la definición establecía que una persona se construye mediante una

negociación narrativa, por lo que aunque las personas no se definen por sus condiciones físicas y pueden cambiar de un soporte material a otro, parece necesario que para que algo o alguien sea una persona su soporte tenga que tener al menos la potencialidad de desarrollar no sólo una narración, sino una vida mental compleja, pues de otro modo no habría negociación posible. Por otro lado, para que alguien o algo sea la misma persona que otra persona pasada, tiene que suceder que su vida mental tenga una continuidad con la anterior, pues de otro modo la negociación que establece su identidad podría romperse, ya fuese porque la persona futura negase ser la persona pasada o porque la sociedad no aceptase dicha identidad al ser el comportamiento de la persona futura radicalmente distinto al de la pasada. Evidentemente, y como ya señaló Williams, su vida mental no será la misma numéricamente aunque sea idéntica cualitativamente a no ser que esté realizada sobre la misma base física. Pero dado que las personas no se definen por tener una misma conciencia (al contrario que los Yoes), sino por ser el objeto de una serie de creencias colectivas, este aspecto no es relevante. Una persona puede sobrevivir a un cambio de soporte físico siempre que tenga la creencia de que va a hacerlo, que la sociedad en la que viva crea que va a hacerlo, y que el nuevo soporte tenga una vida mental que sea continua con la del soporte anterior y dé por tanto lugar a una negociación narrativa coherente con la negociación narrativa en la que estaba inmerso dicho soporte.

Finalmente, aquellos que aún no han sido convencidos por mi argumentación podrían objetar que mi concepto de «persona» no deja de ser como otro tipo de conceptos sociales que también somos y que tampoco se ven afectados por nuestra muerte, como los de «rey» o «presidente del gobierno». Pues en efecto cuando un ser humano que ostenta el título de rey de un país muere, la figura de rey no muere con él, sino que es heredada por su sucesor. Si la analogía fuese válida y el concepto de «persona» que estoy elaborando se pareciese al de «rey», podría ser que a nuestra muerte como seres humanos alguien pudiese heredar nuestra «personidad» del mismo modo que podría heredar nuestro número de pasaporte. Sin embargo un rey no se identifica con su cargo del mismo modo en el que se identifica con el ser humano que es,⁵ ni nosotros nos identificamos con nuestro número de pasaporte en un grado significativo. Nadie pensaría que va a sobrevivir y se alegraría por ello si supiese que alguien va a heredar su número de pasaporte.

5 Merece la pena señalar aquí el caso del Dalái Lama, quien debido a la creencia budista en la reencarnación en principio parece que podría identificarse plenamente con su sucesor.

Sin embargo es importante hacer hincapié en que, a diferencia del heredero de un título nobiliario, una réplica futura mía que fuese mi sucesora como persona no sólo tendría derecho a tener mis posesiones, sino que dado que también tendría una vida mental continua con la mía su comportamiento sería idéntico al que yo –entiéndase que «yo» se refiere aquí a la persona que soy en la actualidad– tendría en su situación. Continuará mis proyectos del mismo modo a como lo habría hecho yo mismo y lo hará reconociéndome a mí como su autor, de modo que podría ser considerado como un mismo centro de acción. En este sentido creo que existen motivos para alegrarnos por nuestra supervivencia como personas que van más allá del hecho de que alguien continuará usando nuestro pasaporte.

Si todo lo dicho hasta ahora es cierto, entonces Tony Soprano tras el experimento sobreviviría como ser humano, como Yo y como persona, pero las tres entidades no compartirían el mismo espacio físico como sucedía anteriormente. De este modo, la muerte de una cualquiera de ellas no afectaría a la supervivencia de las otras dos, por lo que todavía existiría alguien a quien Tony Soprano podría llamar «yo». Sin embargo la muerte de las tres entidades supondría la desaparición completa del mafioso. Por otro lado, su existencia como persona es lo que debería importarle tanto a él como a nosotros en lo que se refiere a temas de responsabilidad moral.

A raíz de este argumento debemos aceptar que la existencia de seres humanos, Yoes y personas es independiente. A pesar de que nosotros somos al mismo tiempo las tres entidades, pueden existir personas que no sean ni seres humanos ni Yoes, seres humanos que no sean ni Yoes ni personas, y Yoes que no sean ni personas ni seres humanos. Incluso en nuestro propio caso, las condiciones de supervivencia de las tres entidades que somos son diferentes, de tal modo que podemos sobrevivir como cualquiera de ellas –y alegrarnos de haber sobrevivido– aunque las otras dos no lo hagan. Sin embargo cuando hablamos de responsabilidad moral sólo nuestra existencia como personas tiene relevancia.

3.2. El materialismo frente al problema de la conciencia

Me parece difícil que las conclusiones que hemos extraído en el apartado anterior sean aceptadas por un materialista en cualquiera de sus versiones. No obstante parece necesario hacer un análisis de la situación, pues el materialismo adopta múltiples formas y podría ser que alguna de ellas nos sirviese de base para sostener nuestras intuiciones acerca de qué entidad nos debe importar en nuestra supervivencia. Para demostrar que

esto no es así, en primer lugar mostraré los problemas que tiene el materialismo para explicar la dimensión fenomenológica de la conciencia. Dado que los Yoes están definidos en virtud de este tipo de propiedades, las dificultades que el materialismo tiene para explicarlas podría ser motivo suficiente como para renunciar al mismo.

Como señalaba en la introducción, tras varias décadas de debate sobre el problema de la conciencia, parece extenderse en la filosofía analítica la creencia de que aceptar el materialismo en última instancia implica comprometerse con algún tipo de reduccionismo cercano al eliminativismo, como ya mostrase Joseph Levine (Levine, 1983, pp. 360-361). Esta postura con respecto al problema de los qualia se puede decir que oscila entre el materialismo eliminativo y el funcionalismo reductivo, en lo que se puede denominar materialismo revisionista (Churchland, 1984, pp. 84-85; Dennett, 1991a, pp. 56-57). Dicha postura se basa en la creencia de que lo que tradicionalmente se han conocido como experiencias conscientes en realidad no son más que estados producidos por la actividad cerebral a los que se les ha otorgado un estatus especial, casi mágico, lo que les da ese aura de irreductibilidad. Sin embargo, una vez depuramos nuestra concepción de la mente para eliminar dichas características sobrenaturales que se le han asociado a la conciencia podemos ver que en realidad lo que anteriormente se habían considerado experiencias conscientes no tienen nada de mágico ni especial, por lo que se pueden identificar con estados cerebrales sin mayor problema.

Este proceso podría describirse como el caso en que lo que antes llamábamos experiencias conscientes ahora se ha descubierto científicamente que son muy distintas a lo que nos imaginábamos, pues son simples estados cerebrales. Pero la diferencia entre lo que nos imaginábamos y lo que la neurociencia nos dice es tan grande que dicha descripción parece forzada. Si alguien nos dijese que ha descubierto la segunda parte de la *Poética* de Aristóteles, y que pese a lo que tradicionalmente se había pensado no fue escrita por el filósofo griego sino por una mujer irlandesa que vivió en el siglo IX, entonces diríamos que no ha descubierto la segunda parte de la *Poética*, sino un libro distinto. Aunque el libro de esta ficticia mujer irlandesa fuese el origen de los rumores sobre la existencia de una parte de la *Poética* dedicada a la comedia, lo que tendríamos entre manos sería tan distinto de lo que esperábamos que publicarlo como si fuese la segunda parte de la obra de Aristóteles sería considerado un fraude. Por interesante que pudiese ser ese libro, debería ser conocido por otro nombre, al mismo tiempo que debería reconocerse que nunca existió una segunda parte de la *Poética*. Del mismo modo, decir que las experiencias conscientes son estados cerebrales es una

descripción poco adecuada, por lo que sería mejor decir que se ha descubierto que las experiencias conscientes no existen, y que su origen mítico estaba en la actividad de nuestras neuronas. En otras palabras, identificar las experiencias fenoménicas con estados cerebrales supone eliminar las primeras del conjunto de cosas que existen realmente, motivo por el que esta postura se denomina «eliminativa».

Esta actitud evidentemente sería compatible con la aceptación de que el Yo es una entidad ficticia, tal y como sostiene Dennett.⁶ Sin embargo, como vimos en la nota al pie 26 del capítulo anterior, esto hace que resulte imposible explicar la importancia de los Yoes en nuestra vida real a no ser que le otorguemos algún tipo de existencia a las ficciones, lo que en última instancia refutaría la premisa materialista que le dio origen. De este modo los partidarios del materialismo eliminativo deberían ofrecer alguna explicación alternativa de la importancia de los Yoes o, por el contrario, aceptar que su punto de partida es auto-contradictorio.⁷

Dejando a un lado momentáneamente el problema de la identidad personal, el eliminativismo ha sido tachado de inverosímil, pues deniega la existencia de lo único que podemos estar seguros: nuestras propias experiencias conscientes. Al mismo tiempo se ha acusado a sus partidarios de dar carpetazo al problema de la conciencia sin enfrentarse al «problema difícil», pues confunden el sentido fenoménico de la palabra con el psicológico, de modo que sus análisis, aunque puedan ser correctos, ni siquiera abordan el tema que prometen explicar (Chalmers, 1996, pp. 16-17). Por mucho que se intente es imposible ofrecer una explicación de los fenómenos subjetivos desde una perspectiva de tercera persona, por lo que en toda reducción de la fenomenología a ciencia objetiva inevitablemente se está perdiendo la perspectiva sobre lo que necesita una explicación.

Quizá el argumento más extendido en la actualidad entre los críticos de las posturas eliminativas es la famosa ficción filosófica de los zombis –aunque también son muy conocidos los experimentos de la nación china de

6 En el siguiente capítulo discuto con mayor detalle la postura ontológica de Dennett, hasta ahora identificada con un materialismo reductivo al uso pero que en realidad se encuentra más cerca de mi posición de lo que podría parecer.

7 Otra opción por la que los eliminativistas pueden decantarse para explicar el Yo es tratar de identificarlo con alguna parte del cerebro, algo que Dennett considera un error categorial (Dennett, 1988, pp. 108-109) pero que no ha dejado de intentarse. Un ejemplo de este tipo de postura se puede encontrar en Paul Churchland, quien parece vincular el Yo con el núcleo intralaminar del tálamo, de modo que considera que una lesión en esa región cerebral provoca su pérdida (Churchland, 1995, p. 306). En cualquier caso, como resulta evidente, este tipo de respuestas renuncian a encontrar una explicación a la dimensión fenomenológica del Yo, por lo que no les prestaré más atención.

Block (1978), el de la habitación china de Searle (1980), o el de Mary la científica del color de Jackson (1982)–, hasta el punto de que para Dennett el debate sobre la naturaleza de la conciencia dentro de la filosofía analítica actual se puede considerar como una discusión entre dos bandos: los que dan credibilidad a la *corazonada zombi* y los que no (*vid.* Dennett, 2001).

Aunque hay varias descripciones del concepto de zombi (Colomina Almiñana, 2008, pp. 162-166), la idea que subyace a todas ellas es la misma. El zombi es un ser capaz de *imitar* el comportamiento humano en todos los sentidos –en algunos casos incluso es físicamente idéntico a nosotros molécula por molécula y sería confirmado como humano por todas las pruebas psicológicas, físicas o médicas que se le pudieran hacer (Chalmers, 1996, p. 133)– pero que sin embargo carece de algo imprescindible para ser considerado un humano: no tiene experiencias conscientes. A pesar de que como buena réplica el zombi diría estar convencido de tener esas experiencias, y de que ninguna prueba realizada desde una perspectiva de tercera persona podría encontrar diferencia alguna con un ser humano normal, el zombi no sería consciente en el sentido en que lo somos nosotros. Si recibe un golpe en la espinilla saltará y gritará exclamando cuánto le duele, pero en realidad no experimentará de forma consciente ningún dolor. Ante la visión del cielo asegurará que es de un precioso color azul, pero lo hará de forma automática, sin que haya habido ninguna experiencia cualitativa del color.

La fuerza del argumento zombi no reside en su aplicación al mundo real. Para el caso no importa si todos los seres humanos que han existido y existirán tienen experiencias conscientes o si la existencia de zombis es física o metafísicamente imposible. Lo que trata de demostrar el argumento es la posibilidad *lógica* de su existencia, el hecho de que podemos imaginar sin caer en una contradicción un mundo en el que aunque todos los eventos descritos por la física fuesen idénticos a los de nuestro mundo, no hubiese ninguna propiedad fenoménica en absoluto. Si imaginar esto es posible entonces el argumento demuestra que la experiencia consciente posee alguna cualidad no reducible a la física. O al menos a la física tal y como es entendida actualmente, por lo que ninguna teoría materialista de la conciencia basada en los conocimientos físicos que poseemos hoy en día podría explicar por completo la conciencia, a pesar de que se puedan encontrar relaciones directas entre todo evento consciente y algún tipo de actividad cerebral.

Para tratar de ofrecer una explicación realista de la conciencia que integre las conclusiones de este argumento en las últimas décadas han aparecido diferentes respuestas que tratan de mostrar que el problema

reside en nuestra concepción ontológica del mundo. Esto es, al haber intentado reducir las experiencias conscientes a la física actual hemos pasado por alto la posibilidad de que la física actual esté incompleta o equivocada, y que no todo pueda reducirse a las propiedades físicas que conocemos hoy en día pero sí a propiedades aún por descubrir. Si esto fuese así aún podríamos encontrar una explicación fisicalista y realista de la conciencia. Lo único que tenemos que hacer es cambiar nuestra opinión sobre qué tipos de propiedades existen según nuestra ontología para poder incluir así las propiedades fenoménicas.

En general podemos encontrar dos tipos de respuesta según el nivel en el que se considere que existen las propiedades fenoménicas, por lo que las podemos clasificar en *emergentistas* –consideran que la conciencia es una propiedad emergente de la materia que aparece cuando se da una organización lo suficientemente compleja de la misma– y *panpsiquistas* –las propiedades fenoménicas son propiedades intrínsecas de la materia física en su nivel fundamental, por lo que todo elemento físico tiene al menos experiencia proto-consciente–. A su vez los partidarios del emergentismo se dividen entre *interaccionistas* –los defensores de que las propiedades fenoménicas interactúan con las físicas– y *epifenomenistas* –los que consideran que el mundo físico está causalmente cerrado, por lo que la conciencia no tiene ninguna influencia sobre él–. Tanto las posturas emergentistas como las panpsiquistas aceptan un dualismo de propiedades, aunque son también compatibles con un dualismo substancialista –la tesis de que la conciencia depende de la existencia de algún tipo de substancia no-física aún no detectada–, aunque no parece haber ningún argumento fuerte a favor de esto último, y de haberlo no queda claro en qué se distinguiría del dualismo cartesiano (Chalmers, 2010, pp. 126-139). En cualquier caso en este momento sólo me interesan las posturas monistas (materialistas) que al mismo tiempo sean realistas.

Las posturas emergentistas son posiblemente las más extendidas junto al materialismo estándar que presenté al principio del capítulo y que podemos denominar «materialismo no tengo ni idea», esto es, la tesis de que la conciencia debe ser física dado que el materialismo es verdadero, aunque no se sabe cuál es la relación entre lo físico y lo fenoménico (Chalmers, 1996, p. 214). El emergentismo surge cuando partiendo de este mismo supuesto de que la conciencia debe ser física se intenta eliminar la incertidumbre acerca de la relación entre lo físico y lo mental. Se trata de una postura que ha existido en los márgenes del debate durante largo tiempo, aunque recientemente ha cobrado un papel más relevante, pues es la postura lógica si uno es materialista y tiene como creencias verdaderas que la conciencia existe y que las partículas fundamentales que constituyen

la realidad carecen de propiedades conscientes (Strawson, 2006, pp. 18-19). Para que estas dos creencias sean compatibles se hace necesario postular que la conciencia debe ser una propiedad emergente de la materia cuando se da un cierto nivel de organización en la misma.

La idea se ilustra generalmente poniendo como ejemplo otras propiedades emergentes como puede ser la liquidez del agua. Las moléculas de H₂O no son en sí mismas líquidas, ya que la liquidez sólo aparece como una propiedad emergente cuando muchas moléculas de agua se unen. Del mismo modo, dicen los emergentistas, puede que las moléculas orgánicas no sean en sí mismas conscientes, pero al unirse de una forma compleja dando lugar a cerebros humanos la conciencia aparece. Sin embargo no hace falta mucho para darse cuenta de que el emergentismo en realidad no soluciona nada por sí mismo, ya que la idea de que la conciencia pueda surgir a partir de la interacción de partículas no-conscientes, sin ofrecer ninguna explicación de cómo es esto posible, es igual de misteriosa que la de que la conciencia es física porque el materialismo es verdadero. La emergencia de la liquidez es explicable por completo a partir de las características físicas de las moléculas individuales, por lo que no es en absoluto sorprendente. Sin embargo la emergencia de la conciencia no sólo no es explicable a día de hoy a partir de la interacción de partículas no-conscientes, sino que ni siquiera podemos imaginar en qué podría consistir una explicación de ese fenómeno (*cf.* Rodríguez González, 2010, pp. 185-188). Defender que esto sucede supone una fe injustificada que nos podría llevar a aceptar cualquier tipo de violación de las leyes de la naturaleza mediante la aparición de propiedades emergentes completamente indeducibles de las propiedades fundamentales de la materia.

Una forma de solventar los problemas asociados con las teorías emergentistas consiste en abandonar la creencia de que las partículas fundamentales que constituyen la realidad carecen por completo de propiedades conscientes. La idea básica es que si admitimos que existe alguna propiedad proto-fenoménica en el nivel fundamental de la materia entonces la emergencia de la conciencia en niveles superiores dejaría de ser ininteligible (Chalmers, 2010, p. 133). Existen dos formas en que esto podría ser cierto. La primera sería admitir que existen partículas fundamentales con propiedades fenoménicas además de partículas fundamentales con propiedades físicas, lo que supone comprometernos con un dualismo substancialista para el que ya hemos dicho que no tenemos grandes argumentos. La segunda opción es considerar que además de contar con propiedades físicas como la masa o la carga eléctrica, las partículas fundamentales poseen también propiedades proto-

fenoménicas (Strawson, 2006, pp. 24-26). La consecuencia lógica de esto es que entonces la experiencia consciente está presente en todos los objetos físicos, motivo por el cual esta postura recibe el nombre de «panpsiquismo».

Las posturas panpsiquistas en un primer momento suenan extravagantes, pero sin embargo no hay nada que nos impida aceptarlas como una respuesta posible. Lo único que requieren es aceptar que las partículas fundamentales de la realidad tienen propiedades proto-fenoménicas intrínsecas, lo cual no es incompatible con la física actual. De hecho la física guarda un silencio llamativo sobre cuál es la naturaleza intrínseca de la materia –cuáles son por ejemplo las propiedades intrínsecas de los quark que dan lugar a la masa– pues define todas sus entidades de forma extrínseca, esto es, en función de las relaciones que establecen con otras entidades (Chalmers, 2010, p. 27).⁸ De este modo el panpsiquismo se constituye como una alternativa realista plenamente compatible con nuestros conocimientos actuales y en principio capaz de solucionar el problema de la conciencia. El único escollo que le falta por superar para cumplir su propósito es explicar cómo las propiedades proto-fenoménicas podrían combinarse para dar lugar a nuestra experiencia consciente (Dennett, 2001, pp. 34-35), pero en cualquier caso están en mejor situación que las posturas emergentistas.

Recapitulando lo dicho hasta el momento, podemos considerar que las posturas materialistas en la actualidad tienen cuatro vertientes principales. La materialista estándar, la eliminativa, la emergentista y la panpsiquista. La primera es una posición preteórica que ya hemos dicho que parece comprometernos con la segunda, mientras que la tercera apenas puede considerarse una postura coherente. De las otras dos opciones, la eliminativa ya hemos visto que no es compatible con una solución apropiada para el problema de la identidad personal, pues por un lado nos obligaría a aceptar que las ficciones existen de algún modo, contradiciendo así su punto de partida de que todo lo que existe es físico, y por otro renuncia a explicar la dimensión fenomenológica de los Yoes. Por su parte el panpsiquismo tiene en su contra el ser una postura altamente inverosímil, y aunque en principio podría ser una solución para el

8 Posteriormente a esta argumentación de Chalmers el CERN anunció el descubrimiento del bosón de Higgs, la partícula que explicaría en el modelo estándar de la física de partículas la existencia de la masa. Sin embargo, otra vez, la física define esta propiedad de la materia mediante la relación extrínseca de unas partículas con otras y no por medio de sus propiedades intrínsecas.

problema de la conciencia, aún tiene que resolver el difícil problema de la combinación.

Las dificultades que todas estas posturas tienen para explicar la conciencia podrían ser consideradas motivos suficientes para abandonar el materialismo sin detenernos más en ello. Sin embargo, todavía me gustaría presentar algunos argumentos más para mostrar que aunque pudiesen solucionar el problema de la conciencia, ni el pansiquismo ni ninguna otra postura materialista podrían constituir una base apropiada para resolver el problema de la identidad personal desde el NPN, pues son incapaces de dar una explicación satisfactoria de nuestras intuiciones acerca de lo que creemos ser como personas y la importancia que su supervivencia tiene para nosotros.

3.3. El materialismo frente al problema de las personas

Como decía al principio del apartado anterior, independientemente del problema de la conciencia me parece difícil que ningún materialista se muestre dispuesto a aceptar mis conclusiones acerca de que cuando nos preocupamos por nuestra supervivencia nos importan en la misma medida las tres entidades que somos. Especialmente con respecto a la importancia que atribuimos a nuestra existencia como personas, pues la mayoría de ellos querrían defender que existe una cierta primacía ontológica del ser humano, cuya existencia en todos los casos debería ser necesaria para la existencia de Yoes y personas. Sin embargo si esto fuese así debería suceder que nuestras condiciones de supervivencia a través del tiempo como los tres tipos de entidades estuviesen relacionadas, de tal modo que, como mínimo, la muerte del ser humano implicase necesariamente la muerte del Yo y de la persona. Este sería el caso por ejemplo si el concepto de «persona» fuese parecido a otro concepto social como es el de «adolescente», de modo que se aplicase a las etapas de nuestra vida en tanto seres humanos en las que tenemos unas ciertas capacidades mentales (Parfit, 2012, pp. 8-9). Si esto fuese cierto, nuestra muerte como seres humanos debería implicar nuestra muerte como personas del mismo modo que la muerte de un ser humano de quince años de edad implica la muerte del adolescente que también es; al mismo tiempo nuestra desaparición como personas podría darse aunque siguiésemos viviendo en tanto seres humanos, igual que desaparecemos como adolescentes al cumplir los veinte años. Sin embargo, como vimos en el primer apartado, esto no es así. Nuestra muerte como seres humanos puede darse sin afectar a nuestra existencia como Yoes o como personas.

Uno de los motivos por los que considero que la mayoría de los materialistas se sentirán inclinados a otorgar al ser humano una primacía ontológica sobre las personas es que buena parte de ellos consideran la realidad social como una mera construcción ficcional, una forma que tienen los seres humanos de comprender el mundo imponiéndole falsas creencias y que por tanto no merece mayor atención. En este sentido, para muchos materialistas las personas que hemos definido no existirían más que en el mismo sentido en que pueden existir otros personajes de ficción, como por ejemplo el mafioso protagonista de la serie *Los Soprano* (David Chase, cr., 1999-2007). Sin duda un personaje como Tony Soprano puede «sobrevivir» cambiando de soporte, pues puede ser interpretado por un actor diferente. Y sin duda también un personaje puede morir sin que el actor que lo interpreta muera con él. Pero cuando nos preocupamos por el mundo real, lo que verdaderamente importa es la supervivencia de James Gandolfini, el actor, no la de su personaje. Que Tony Soprano muera es irrelevante en el mundo real.

Un segundo motivo es que otra buena cantidad de materialistas considera que la realidad social tiene una existencia derivada, dependiente de los hechos brutos que componen la realidad física (*vid.* Searle, 1995). De este modo aunque aceptan que la realidad social existe, parece que para ellos no existe *tan realmente* como existe la realidad física subyacente. Sin duda en cierto sentido existen los países, existe España por ejemplo. Pero no existe tan realmente como el territorio físico y los seres humanos que lo conforman. Y si nos preocupamos por su supervivencia, sin duda es menos grave que España se divida en distintos países y sus habitantes adquieran nuevas nacionalidades, a que una bomba nuclear mate a todos sus habitantes y destruya por completo su territorio haciéndolo desaparecer bajo el mar. Lo segundo es una verdadera pérdida, lo primero un simple cambio en el modo en que los seres humanos dividen arbitrariamente el mundo y regulan sus relaciones. Pero si la analogía es válida, lo que defiende la tesis de que nos deben importar en igual medida la supervivencia de las tres entidades que somos es que nos debe parecer igual de grave la disgregación de España que su desaparición debido a una bomba nuclear. Y que la muerte de Tony Soprano es tan importante como la de James Gandolfini.

Ante este panorama, no creo que muchos materialistas estén dispuestos a aceptar que nuestra supervivencia como personas es tan importante como nuestra supervivencia como seres humanos o como Yoes. Sin embargo se me ocurren dos formas en que podrían hacerlo, por lo que trataré de armar dos casos en mi contra que son compatibles con el eliminativismo y el pansiquismo respectivamente.

En primer lugar, independientemente de su incapacidad para explicar la dimensión fenomenológica de nuestra vida, los eliminativistas aceptan la existencia de estados mentales que son reducibles a estados neuronales. Por otro lado, las creencias colectivas que constituyen a las personas podrían ser entendidas en este marco como el conjunto de las creencias al respecto de todos los individuos de una sociedad. De este modo, y dado que las creencias no son más que estados mentales reducibles a estados neuronales, un eliminativista podría sostener que las personas existen realmente, pero que tienen una base física insospechada: las neuronas de todos los componentes de una sociedad en la que están codificadas las creencias acerca de su existencia. De este modo, el hecho de que las condiciones de supervivencia de seres humanos, Yoes y personas sean independientes no es ninguna sorpresa, pues las tres entidades tienen bases físicas distintas.⁹

Sin embargo esta postura presenta varios problemas. Primero, la tesis de que podemos identificarnos en la misma medida con las tres entidades parece difícil de justificar en este marco, pues suena bizarro decir que lo que me importa en mi supervivencia (como persona) es la supervivencia de las neuronas de mi vecino, pues en cierto sentido *soy* esas neuronas. Segundo, esta postura ignora el hecho de que no es necesario que las creencias colectivas que dan lugar a la realidad social sean conocidas explícitamente por nadie. Poniendo como ejemplo un hecho social como es el lenguaje, la mayoría de los hablantes ignoran por completo las reglas de uso del mismo, pues se limitan a hablarlo, no a pensar sobre él. De hecho las reglas no están *ahí* para ser descubiertas por los estudiosos, sino que son elaboradas por los filólogos para comprender un fenómeno que se da al margen de toda regla explícita. De este modo puede darse un hecho social sin necesidad de que las creencias colectivas que lo constituyen estén codificadas en ningún lugar. Al mismo tiempo las personas pueden albergar todo tipo de falsas creencias acerca de hechos sociales sin que esto afecte al hecho social mismo. Por ejemplo durante siglos en Europa se creía que los reyes eran reyes por derecho divino, sin que esa falsa creencia afectase al hecho de que eran reyes por otros motivos. Y tercero, esta postura reduccionista tiene una consecuencia desagradable, y es que no puede justificar cuál es la diferencia entre que alguien sobreviva como persona gracias a una réplica que sea socialmente considerada como tal, y que alguien sobreviva gracias a que todo el mundo mantenga la falsa creencia de que ha sobrevivido. Si lo que yo soy como persona son las neuronas de mis vecinos, lo que importa es que esas neuronas sigan funcionando del modo en que lo hacían mientras yo existía. Que lo hagan

9 Agradezco a Miguel Ángel Sebastián sus comentarios sobre este tema.

debido a la existencia de una réplica o debido a la existencia de un rumor no parece especialmente relevante. Siendo así, considero que nuestras intuiciones acerca de aquello que nos debe importar en nuestra supervivencia son incompatibles con una postura reduccionista del tipo que he descrito.

En segundo lugar, un panpsiquista o un emergentista –como vimos en cierto sentido los panpsiquistas son emergentistas– podría sostener lo que Lynne Rudder Baker llama la tesis de la «constitución sin identidad», y que podemos entender mejor a partir de un ejemplo (Baker, 2000, pp. 27-58). Consideremos una estatua –por ejemplo el *David* de Miguel Ángel– y el trozo de mármol del que está hecha –llamémoslo Trozo–. La mayoría de la filósofos piensan que la estatua es idéntica a dicho trozo de mármol. Sin embargo Baker señala que la relación entre la estatua y el trozo de mármol no es de identidad, sino de constitución. *David* no es idéntico a Trozo ni tampoco es un objeto distinto, sino que está constituido por Trozo. *David* no está compuesto por nada aparte de Trozo, y estrictamente hablando no existe nada en *David* aparte de las moléculas de Trozo. Sin embargo *David* posee características que Trozo no posee, por lo que merece ser considerado como un objeto distinto. Por ejemplo, una de las propiedades modales de Trozo es que podría seguir existiendo en un mundo en el que no existiese el arte, mientras que *David*, al ser esencialmente una obra de arte, no podría hacerlo. Por tanto *David* posee una propiedad, la de ser una estatua en cualquier mundo en el que exista, que Trozo no posee. Sin embargo Trozo y *David* no son dos objetos independientes, sino que uno constituye al otro. En líneas generales la tesis de la constitución sin identidad defiende que cuando ciertas entidades con unas propiedades determinadas se encuentran en las circunstancias apropiadas, aparecen nuevas entidades que poseen nuevas propiedades. Cuando tenemos un trozo de mármol con unas propiedades concretas en un mundo en el que existe el arte, aparece un nuevo objeto con propiedades que el trozo de mármol no posee: una estatua. O más exactamente, aparece un objeto que posee de manera esencial propiedades que el objeto que lo constituye posee de forma contingente. Trozo es una estatua en nuestro mundo, pero podría dejar de serlo, por lo que ser una estatua es una propiedad contingente de Trozo. Por su parte *David* es una estatua de forma esencial, pues no podría dejar de serlo sin dejar de existir.

Si aplicamos esta idea a los conceptos de «ser humano», «Yo» y «persona»,¹⁰ podríamos decir que la base física de lo que somos es un ser

10 Es importante señalar que aquí me distancio de la postura de Baker, que no distingue entre Yoes y personas. Un desarrollo de su tesis de la constitución sin identidad que sí trata de dar cabida a la dimensión social de nuestra existencia es la que realiza Robert Wilson (2005).

humano. Pero que dicho ser humano, que posee contingentemente autoconciencia, en virtud de poseerla constituye una nueva entidad, un Yo, que la tiene de forma esencial. Un ser humano podría dejar de ser autoconsciente sin morir (por ejemplo quedándose en coma), pero un Yo no podría dejar de serlo sin desaparecer. Por otro lado, en virtud de existir en un mundo en el que nuestras relaciones sociales determinan que somos personas, un Yo constituye una nueva entidad, una persona, que posee esencialmente una serie de propiedades sociales que el Yo posee sólo de forma contingente. Un Yo podría sobrevivir a la desaparición de nuestra sociedad, pero una persona no podría. Por último, desde esta postura se podría defender que las personas tienen esencialmente una serie de propiedades sociales pero sólo contingentemente una serie de propiedades físicas y mentales, por lo que podría suceder que una persona sobreviviese a la pérdida de sus propiedades físicas y mentales originales si conservase las mismas propiedades sociales. De este modo en principio parece que los emergentistas podrían encontrarse en mejor situación que los eliminativistas para explicar nuestras intuiciones acerca de qué entidades nos deben importar cuando nos preocupamos por nuestra supervivencia.

¿Cuál es entonces el problema de esta postura? Si la tesis de la constitución sin identidad decide otorgar una categoría ontológica nueva a las entidades constituidas en lugar de decir que las entidades constituyentes han adquirido una propiedad contingente, es porque considera que las propiedades en cuestión son de una significación mayor que el resto de propiedades que ya tenían las entidades constituyentes (*cfr.* Baker, 2000, p. 33). De este modo, la propiedad de tener autoconciencia se considera más importante que la de tener un hígado, pongamos por caso, por lo que prefiere otorgar a los seres humanos que tienen autoconciencia (además de hígados) un nuevo estatus que los diferencie de otros animales que también tienen hígados pero no son autoconscientes; y la propiedad de poder establecer una relación amorosa se considera más importante que la de poder sentir hambre, por lo que decide otorgarle un nuevo estatus a los Yoes capaces de tener relaciones amorosas que los diferencie de los que únicamente pueden sentir hambre y otras cuestiones similares. Siendo así, lo que resulta difícil justificar en este marco no es que nos preocupemos por nuestra supervivencia como personas, sino que lo hagamos por la supervivencia de los Yoes y los seres humanos que nos constituyen pero que no tienen las propiedades que nos hacen especiales y nos diferencian de otros seres vivos.¹¹ Dado que considero que en nuestra

11 En este sentido Parfit recientemente ha defendido la bizarra tesis de que en realidad no somos seres humanos, queriendo con ello significar que no nos debe importar la supervivencia de los que somos como seres humanos (*vid.* Parfit, 2012).

supervivencia nos deben importar en igual medida las tres entidades que somos y no únicamente las personas, concluyo que la tesis de la constitución sin identidad tampoco nos permite hacer compatibles el materialismo con las tesis que estoy defendiendo.

Alguien aún podría argumentar que existen otras alternativas materialistas que podrían justificar la tesis de que cuando nos preocupamos por nuestra supervivencia nos importan las tres entidades que se distinguen en el NPN. Lo único que puedo decir es que no las conozco. Y de haberlas, no sé si su contenido intuitivo sería mayor que el de la alternativa que presentaré en el próximo capítulo. Ante esta situación sólo me queda por decir que renunciar al materialismo es el primer paso en la búsqueda de la mejor solución posible al problema de la identidad personal.

CAPÍTULO 4

UNA PROPUESTA ANTI-REALISTA

En el capítulo anterior mostré que es imposible justificar nuestras intuiciones acerca de lo que nos debe importar en nuestra supervivencia si mantenemos un marco materialista, por lo que abogué por la búsqueda de un nuevo marco. Sin embargo no es mi intención defender que el problema de la identidad personal deba resolverse mediante algún tipo de dualismo. Muy al contrario, mi interés no reside en negar la validez de la teoría atómica de la materia o de la teoría de la evolución de las especies. Un proyecto así sin duda estaría también abocado al fracaso. Lo que pretendo es negar la presuposición de que si estas dos teorías son ciertas, entonces todo debe poder reducirse a una explicación física. Existen aspectos que el materialismo es incapaz de explicar –como la experiencia consciente o nuestras intuiciones acerca de lo que nos debe importar en nuestra supervivencia– y explicaciones perfectamente válidas del mundo que sin embargo no se pueden reducir a una explicación materialista.

En este sentido lo que me gustaría defender en este capítulo es que para resolver de una forma satisfactoria el problema de la identidad personal no es necesario renunciar al materialismo en sí, sino a la idea de que el materialismo debe ser la única explicación de todo lo que existe. El materialismo ofrece una visión de la realidad bastante útil para comprender algunos aspectos de nuestra vida, pero es incapaz de explicar otros, por lo que debe convivir con otras explicaciones igual de válidas (*cf.* Rodríguez González, 2010). Es decir, lo que sostendré es que para justificar nuestras intuiciones acerca de lo que nos debe importar en

nuestra supervivencia no es necesario un cambio de nuestros presupuestos ontológicos –además de objetos físicos existen *realmente* objetos fenoménicos y sociales–, sino un cambio de nuestra actitud epistemológica –suponer que existen entidades físicas es útil para explicar algunas cosas y suponer que existen entidades fenoménicas y sociales es útil para explicar otras, pero no tenemos motivos para asegurar la existencia *real* de ninguno de los tres tipos de entidades–. En otras palabras, y volviendo a la descripción del materialismo estándar que di al principio del capítulo anterior, no se trata de cuestionar la validez de la física o la biología actual, sino la de la teoría de la verdad como correspondencia que frecuentemente acompaña a ambas y que nos lleva a pensar que debemos encontrar una explicación satisfactoria en términos físicos de cualquier tipo de entidad para poder asegurar que existe realmente (*vid.* García Ruiz, 2009b, pp. 213-214). Aunque el materialismo sea verdadero, eso no implica que otras teorías que sean contradictorias con él no puedan serlo también, pues ninguna de ellas adquiere su validez de su correspondencia con una realidad externa ontológicamente objetiva.

La defensa de este aspecto me llevará a una serie de autores que, tras la crisis del conductismo lógico a principios de los años sesenta, atacaron la actitud realista en la filosofía de la ciencia –esto es, la creencia de que la ciencia describe una realidad objetiva externa a nuestras percepciones de la misma– inspirándose en la historia y la sociología de la disciplina, además de en el Wittgenstein de las *Investigaciones Filosóficas* (Diéguez Lucena, 1998, pp. 13-14). Autores como Willard Quine, Thomas Kuhn, Paul Feyerabend, Richard Rorty o Nelson Goodman coincidían en el hecho de que el lenguaje que utilizamos para describir la realidad, lejos de ser neutro, supone necesariamente la imposición de una serie de esquemas y valores subjetivos, al mismo tiempo que la actitud del científico hacia su objeto de estudio nunca es ajena a las condiciones políticas, sociales o psicológicas que le afectan. De este modo la ciencia avanza no gracias a un método racional y objetivo que le permite descubrir cómo es la realidad descartando las teorías erróneas en favor de otras más cercanas a la verdad, sino que toda teoría científica es un conjunto de postulados absolutamente subjetivos a pesar de su utilidad para predecir determinados aspectos del mundo que nos rodea. En última instancia, y dada la imposibilidad de separar los supuestos hechos explicados por una teoría y el lenguaje utilizado para describirlos, la subjetividad del conocimiento científico implica que toda teoría es siempre inconmensurable con cualquier otra, pues no es cierto que las teorías rivales aborden los mismos problemas, sino que cada teoría define sus propios problemas que son inexpresables en el lenguaje de sus rivales. Siendo así debemos renunciar

a la idea de que el conocimiento científico se esté acercando a la verdad en el sentido de ser cada vez una descripción más fiel de la realidad objetiva y aceptar en su lugar que toda teoría presupone o construye un mundo ontológicamente subjetivo distinto e irreducible a los de otras teorías.

Una vez contemos con estas herramientas, estaremos en disposición de enfrentar de nuevo el problema de la identidad personal, mostrando cómo podemos justificar nuestras intuiciones sobre lo que nos debe importar en nuestra supervivencia si suponemos que las tres entidades que se distinguen en el NPN se refieren a nosotros bajo distintas descripciones del mundo. Sin embargo esto sólo quedará apuntado al final de este capítulo para poder ser desarrollado en el próximo. Antes de eso, dedicaré unas breves páginas a analizar cómo podemos abordar en este marco la diferencia entre mundos válidos e inválidos, reales y ficticios y científicos y artísticos, cuestiones todas ellas relevantes para las teorías de la identidad narrativa.

Pero antes de empezar a exponer más en detalle las críticas al realismo y la alternativa que propongo creo que es importante aclarar un aspecto que hace que las posturas anti-realistas no sean muy bienvenidas en determinados ambientes. Calificarse a uno mismo de anti-realista, constructivista, pragmatista, instrumentalista o postmoderno es visto a menudo en el ámbito de la filosofía analítica como una afirmación a favor de algo parecido a que si realmente lo creemos posible podremos atravesar una pared con sólo intentarlo, y que cuando cerramos los ojos el mundo (literalmente) desaparece. Se piensa en el anti-realismo como en una especie de idealismo radical, o como en una versión paródica de *Matrix* (Larry y Andy Wachowski, dirs., 1999) en la que la realidad se pliega a nuestros deseos. Sin embargo, aunque ha habido defensores de posturas de este tipo, sostener una actitud anti-realista en filosofía de la ciencia no implica defender que no existe una realidad ontológicamente objetiva, del mismo modo que afirmar que algo es construido no implica que lo podamos construir como queramos (*cf.* Goodman, 1978, p. 131). El anti-realismo en su forma más básica no es una postura ontológica acerca de cómo (no) es la realidad. Es una tesis epistemológica acerca de la imposibilidad de tener un acceso a la misma que no esté mediado por nuestra subjetividad. En otras palabras, es la aceptación de que aunque pueda existir un punto de vista de Dios sobre la realidad nosotros nunca podremos acceder a él, por lo que la ciencia –y por «ciencia» entiendo toda forma de conocimiento científico, incluida la filosofía– debe dejar de considerarse como un modo de conocer el mundo tal y como es para ser interpretada como una manera más o menos útil de explicar y predecir su comportamiento. En este sentido el anti-realismo que sostengo en estas

páginas concede que existe una realidad ontológicamente objetiva –en adelante REALIDAD– que es incognoscible pero cuya interpretación por nuestra parte da lugar a la postulación de la existencia de una serie de realidades ontológicamente subjetivas. De este modo aunque no podamos decir que existe objetivamente delante de nosotros una pared que no podemos atravesar, toda interpretación que hagamos del mundo, para ser válida, debe dar una explicación satisfactoria de la misma «parte» del mundo REAL que nos lleva a nosotros a decir que hay una pared que no podemos atravesar, independientemente de cómo la conceptualice subjetivamente y de que las distintas interpretaciones válidas que se den de la misma puedan ser incompatibles o inconmensurables entre sí. En cualquier caso esto requerirá una explicación más detallada.

Por último antes de terminar esta introducción es importante señalar lo vasto de la tarea que se proponen estas páginas. Analizar en profundidad las cuestiones aquí abordadas necesitaría por sí misma cientos de páginas, y es una labor que excede por mucho los objetivos propuestos en este trabajo. Por este motivo debe tenerse en cuenta que en ningún caso trato de analizar de forma exhaustiva los distintos argumentos que se han dado a favor y en contra del materialismo o del realismo. En su lugar únicamente pretendo justificar los motivos que me llevan a defender mi postura y contextualizarla dentro de una corriente de pensamiento encabezada por autores ampliamente reconocidos en el panorama filosófico, dejando para otro momento la defensa última de sus presupuestos.

4.1. Críticas al realismo

El realismo parte de dos tesis que considera incuestionables. Por un lado, la presuposición de que existen hechos brutos o fundamentales que podemos conocer y que son independientes de nuestra concepción del mundo. Por otro, el consenso de que existe algo llamado «método científico» que permite ir de forma objetiva desde dichos hechos a la verdadera explicación de los mismos. La fuerza de esta segunda tesis se basa en la aceptación de que la primera es verdadera, por lo que si pudiésemos demostrar que no existen hechos brutos la validez del método científico desaparecería. Sin embargo existen argumentos convincentes en contra de las dos tesis, como veremos a continuación.

Según defienden los realistas, un hecho es bruto cuando es totalmente separable de las distintas descripciones que podamos hacer de él, por lo que toda explicación del mundo debe reconocerlo como lo que es. Para un realista lo que tenemos delante de nosotros es una pared, aunque

distintas personas puedan llamarlo «muro», «wall», «wand» o «vegg». Independientemente de la palabra que utilicemos, para un realista el hecho referido es el mismo y es incuestionable. Gracias a la existencia de este tipo de hechos los realistas aseguran que nuestro conocimiento está basado en una correspondencia con la REALIDAD y que por tanto no se ve influido por nuestro punto de vista subjetivo.

Los hechos que son considerados brutos o fundamentales según los distintos autores varían enormemente incluso dentro de los materialistas estándar, pues mientras los nihilistas sostienen que sólo existen las partículas fundamentales de las que está compuesta la realidad, los universalistas afirman que también existen todos los agregados posibles de estas partículas, mientras que los monistas aseguran que únicamente existe un objeto que es el universo en su totalidad (*vid.* Schaffer, 2007). De este modo un nihilista afirmaría que objetivamente no existen las paredes, las mesas, las sillas o las manzanas, sino sólo los átomos, los quarks o las cuerdas de las que en última instancia estén compuestas. Por su parte un universalista no sólo concedería que existen objetos como los átomos, las sillas y las mesas, sino también los objetos compuestos por agregados aleatorios de sillas y mesas o de peras y manzanas. Sin embargo podemos decir que la postura realista dominante es una ontología de «sentido común», que establece que además de las partículas fundamentales de las que esté compuesta la realidad, también existen algunos de sus agregados a los que se puede conceder la categoría de objeto fácilmente. En este sentido la opinión más extendida es que además de cuerdas existen los quarks, los átomos, las moléculas, las paredes, las sillas, los seres humanos, etc., pero no los agregados aleatorios de una silla y una mesa, un plato y unos pantalones o una pera y un platillo volante. Un ejemplo de este tipo de postura podría ser la de John Searle, para quien los hechos brutos se refieren a objetos muy variados y pueden ser tanto del tipo «éste átomo es de hidrógeno» como otros más complejos como «esa montaña está cubierta de nieve» (Searle, 1995, pp. 162-163). Para un realista como Searle, el que una montaña esté cubierta de nieve es un hecho tan objetivo e independiente de nuestras opiniones como el que las moléculas de agua están compuestas por dos átomos de hidrógeno y uno de oxígeno o el que la Tierra gira alrededor del Sol.

Pero precisamente el movimiento de traslación de la Tierra alrededor del Sol es uno de los ejemplos más utilizados por los anti-realistas para negar la existencia de hechos brutos. Una teoría astronómica puede considerar que el enunciado «la Tierra gira alrededor del Sol» refleja un hecho dado de antemano, ante el que nuestras opiniones son irrelevantes y que por tanto nos permite construir sobre él una ciencia que hable de la

REALIDAD. Pero sin embargo todo hecho es siempre relativo a un marco conceptual previo que lo hace inteligible, en este caso el heliocentrismo. Para describir cualquier tipo de movimiento hemos de establecer un sistema de referencia, un punto fijo que nos permita valorar cómo varían las posiciones de los objetos a lo largo del tiempo. En el caso del Sistema Solar este punto fijo se establece normalmente en el Sol, pero en realidad no hay ningún motivo que nos obligue a hacerlo así. Podemos establecerlo igualmente en la Tierra y decir que es el Sol el que gira a su alrededor, ya que no tenemos razones para decir que el Sol está quieto –¿que está quieto respecto a qué?– a no ser la mayor simplicidad que se obtiene así al describir los movimientos del Sistema Solar. Esto es, nuestra decisión habitual de situar el eje de nuestro sistema de referencias en el Sol se basa en la utilidad que podemos extraer de describir así el mundo, no en una mayor correspondencia con la REALIDAD. Sin duda las posiciones relativas del Sol y de la Tierra cambian a lo largo del tiempo, pero para poder describir dichos cambios hemos de decidir cuál va a ser nuestro sistema de referencia. Como tomar dicha decisión es inevitable, pero no existe una necesidad lógica que nos fuerce a elegir un sistema en concreto, hemos de aceptar que nuestra decisión será convencional, esto es, artificial. Pero si el marco que elegimos es convencional, el movimiento que describamos también lo será, de modo que el supuesto hecho que vinculaba nuestra teoría astronómica con la REALIDAD –que la Tierra gira alrededor del Sol– no es un hecho objetivo dado de antemano, sino que se trata de un hecho construido a partir de nuestra decisión arbitraria de utilizar un marco de referencia en particular (Goodman, 1978, pp. 154-158). Pero no sólo el movimiento, sino que la misma existencia de la Tierra, el Sol o cualquier otro objeto supone de forma análoga un proceso de individuación basado en decisiones convencionales. Los hechos están siempre cargados de teoría. En consecuencia si todos los hechos dependen de un marco de referencia previo entonces debemos aceptar que los hechos brutos que supuestamente fundamentan el realismo adquieren la categoría de mito –el mito de lo dado (*vid.* Sellars, 1963)– y deben eliminarse de toda concepción del mundo que se pretenda plenamente racional.¹

1 Para Anthony Rudd el principal problema de los anti-narrativistas es precisamente que parecen aceptar el mito de lo dado, de modo que piensan que entender nuestra vida en forma de narración supone una falsificación de los verdaderos hechos, que son los que deberían decidir sobre nuestra identidad. Sin embargo señala Rudd que los hechos no son separables de las narraciones mediante las que los comprendemos, pues la narración constituye el propio hecho y por tanto no puede falsificarlo. En cualquier caso, aunque carecer del punto de vista de Dios nos impide conocer lo que sería algo así como la

Por otro lado, aunque fuese posible encontrar un sistema de referencia que no fuese convencional –algo así como el sistema de referencia que utilizaría Dios–, los hechos implicados por dicho sistema que nosotros podríamos conocer seguirían siendo necesariamente contruidos desde nuestra subjetividad, por lo que ni siquiera así podríamos hablar de hechos brutos. A pesar de los éxitos explicativos que pudiese tener una ciencia contruida bajo ese hipotético marco –digamos alguna interpretación de la mecánica cuántica–, lo único que podríamos saber es que describir la REALIDAD postulando la existencia de átomos, electrones y quarks sería útil para explicar una serie de fenómenos que en última instancia sólo podríamos observar en primera persona y conceptualizar gracias a nuestro lenguaje. Es decir, aunque no podríamos observar el movimiento de las partículas subatómicas que nuestro marco definiría, sí podríamos observar sus supuestos efectos en nuestros aparatos de medición, y es para interpretar las lecturas que darían nuestros aparatos de medición que sería útil postular la existencia de dichas partículas. Al mismo tiempo cualquier aspecto de la REALIDAD que no tuviese efecto en los fenómenos a los que somos sensibles quedaría ineludiblemente fuera de nuestra explicación. En este sentido toda la ciencia está contruida en el fondo, y de forma necesaria, desde una perspectiva de primera persona. Aunque pretenda hablar de la REALIDAD, se trata de una ciencia subjetiva. Somos incapaces de salir de nuestra perspectiva de primera persona porque para comprender el mundo debemos analizarlo desde nuestras percepciones y nuestros conceptos y no podemos comparar nuestros conceptos científicos con la REALIDAD sin conceptualizar (Quine, 1950, p. 129).

Por último, en cuanto a la validez del método científico y su capacidad para ofrecer explicaciones objetivas, podemos señalar que aunque tanto el marco de referencia como los hechos que se incluyen en dicho marco fuesen ciertamente objetivos y por tanto pudiésemos considerarlos brutos o fundamentales, los hechos por sí solos nunca son lo suficientemente fuertes como para hacernos aceptar o rechazar una teoría científica de un modo racional. Aunque una teoría partiese de hechos brutos, toda interpretación supone imponer un esquema artificial sobre los mismos pues los hechos pueden ser interpretados de diversas formas y a menudo las decisiones que toma la comunidad científica entre dos teorías rivales se basan en criterios subjetivos no-científicos, como pueden ser los principios religiosos, las presiones de determinados grupos de poder o un simple procedimiento democrático: la teoría que consiga un mayor número de

«verdadera y definitiva» narración de nuestra vida, esto no implica que no podamos tener narraciones mejores y peores de una misma situación (Rudd, 2007, p. 66).

adeptos se convierte en la portadora de la verdad (Feyerabend, 1975, pp. 16-18). Así lo que hoy es considerado pseudociencia, mañana puede integrarse dentro de la academia sin que medie ningún cambio sustancial, como por ejemplo puso de evidencia la concesión del Premio Nobel de Química en 2011 a Daniel Schechtman por sus investigaciones sobre los cuasicristales cuando años antes el dos veces Premio Nobel Linus Pauling (de Química en 1954 y de la Paz en 1962) llegó a decir en tono despectivo sobre Schechtman que sólo decía tonterías y que no existían los cuasicristales, sino los cuasicientíficos.² De este modo debemos aceptar que la ciencia avanza no gracias al uso de un método infalible que nos lleva de los hechos a su explicación, sino que lo hace construyendo hipótesis plausibles que son validadas no de forma objetiva –es decir, contrastándolas con la REALIDAD o con el punto de vista de Dios– sino intersubjetiva. Aunque todos sentimos la tentación de pensar que una explicación mítica de un hecho que atribuya por ejemplo el funcionamiento de una máquina a la existencia de un espíritu en su interior es menos acertada que una explicación científica que hable de válvulas, niveles de presión y coeficientes de rozamiento, lo cierto es que la única diferencia entre ambas es su utilidad a la hora de predecir el comportamiento de la máquina o de repararla si se avería, no su adecuación a la REALIDAD.³ Y aunque podemos suponer que una

2 Tras la concesión del premio el propio Schechtman lo comentaba en algunas entrevistas. Véase por ejemplo la siguiente recogida en el blog «Blogoterráqueo» del periodista Sal Emergui, perteneciente a el diario *El Mundo*, el día 10 de octubre de 2011 bajo el título «El Nobel de Química se sincera y aconseja»: <http://www.elmundo.es/blogs/elmundo/blogoterraqueo/2011/10/10/el-nobel-de-quimica-se-sincera-y.html>

3 Véase el siguiente fragmento de la novela *El corazón de las tinieblas* de Joseph Conrad:

Y de cuando en cuando tenía que ocuparme del salvaje que trabajaba de fogonero. Era un ejemplar perfeccionado; podía encender una caldera vertical. [...]. Unos cuantos meses de preparación habían sido suficientes para aquel muchacho realmente estupendo. Escudriñaba el manómetro de vapor y el indicador del nivel de agua con un evidente esfuerzo de intrepidez; [...]. Hubiera debido estar dando palmas y brincos en la orilla, en lugar de lo cual se esforzaba en su trabajo, presa de un extraño maleficio, lleno de un conocimiento provechoso. Era útil porque había sido instruido y lo que sabía era esto: que cuando el agua desapareciera de aquella cosa transparente, el espíritu maligno que se hallaba dentro de la caldera se pondría furioso por causa de la enormidad de su sed, y se tomaría una terrible venganza. Y así sudaba, encendía y miraba el cristal temerosamente (con un amuleto improvisado, hecho con trapos, atado a su brazo, y un trozo de hueso pulimentado del tamaño de un reloj, que le

explicación útil está más cerca de la verdad que una inútil, no podemos comprobar si efectivamente lo está o si es útil simplemente por casualidad –los defensores del realismo científico deberían recordar la gran exactitud del modelo ptolemaico para predecir los movimientos astrales–.

4.2. Pragmatismo y el problema de la convergencia

A menudo se intenta extraer como conclusión de las críticas anteriores que no existe una REALIDAD objetiva, sino sólo intentos de explicaciones de la misma. Que no hay un mundo, sino sólo distintas descripciones del mundo (*vid.* Goodman, 1978). Sin embargo creo que es una necesidad lógica para que puedan existir descripciones que haya algo que describir, aunque ese algo tenga el estatus de mito inalcanzable (*cf.* Searle, 1995, p. 174). Se hace necesario distinguir entonces entre la REALIDAD objetiva y nuestros intentos de descripciones de la misma. Más aún cuando dichos intentos, como hemos visto, no pueden lograr su objetivo. En este sentido creo que la mejor propuesta que se puede hacer para dar una explicación que incorpore las críticas al realismo es admitir que frente a una REALIDAD externa ontológicamente objetiva pero incognoscible que es la fuente de nuestras sensaciones, nuestras diversas explicaciones lo que hacen es postular la existencia de distintas realidades también externas aunque ontológicamente subjetivas. Sin embargo, frente a lo que un realista se sentiría tentado a decir, dichas realidades postuladas no deben tener como objetivo ser lo más parecidas posible a la REALIDAD objetiva aunque tengan en ella su origen, ya que como hemos visto carecemos de un método –cuya existencia, repetimos, es lógicamente imposible– para comparar nuestras conceptualizaciones del mundo con la REALIDAD sin conceptualizar. De este modo, aunque podría ser que por azar la realidad postulada por nuestra ciencia fuese cada vez más parecida a la REALIDAD, si finalmente consiguiésemos hacerlas coincidir no podríamos saber que lo hemos hecho, por lo que el objetivo de la ciencia no debe ser realista, sino pragmático:

Afirmo, por tanto, que carece de sentido preguntarse por la corrección absoluta de un esquema conceptual en tanto que espejo de la realidad. Nuestro criterio para apreciar cambios básicos en un esquema conceptual no debe ser un

atravesaba horizontalmente el labio inferior), mientras las orillas cubiertas de árboles pasaban deslizándose ante nosotros, el ruidito quedaba atrás, interminables millas de silencio se sucedían, y nosotros seguíamos arrastrándonos hacia Kurtz. (Conrad, 1899, p. 186)

criterio realista de correspondencia con la realidad, sino un criterio pragmático. Los conceptos son lenguaje, y la finalidad de los conceptos y del lenguaje es la eficacia en la comunicación y en la predicción. Tal es el deber último del lenguaje, de la ciencia y de la filosofía, y en relación con ese deber debe apreciarse en última instancia un esquema conceptual. (Quine, 1950, p. 129)

Esto es así porque si no existe un criterio para comprobar el valor de verdad de nuestras afirmaciones –esto es, su adecuación a la REALIDAD– entonces el concepto mismo de verdad carece de sentido. Decir que la REALIDAD está compuesta de átomos es una sentencia vacía si no podemos acceder a la REALIDAD desde el punto de vista de Dios para cotejar su validez. Como ejemplo podemos imaginarnos a una civilización alienígena completamente ciega al color que, por motivos religiosos, tiene un interés desmesurado por saber cuál es el color del cielo visto desde la Tierra. Una de sus naves de exploración podría abducir a dos humanos y, tras descifrar su lenguaje, preguntarles por el color del cielo. Uno de ellos, temeroso de lo que podrían hacerle si les mintiese en lo más mínimo, podría decirles que es azul. Sin embargo el otro podría decidir tomárselo a broma y decirles que es naranja. Sus respuestas podrían ser tomadas en serio por los alienígenas, que se sumirían en una profunda discusión sobre cuál de las dos respuestas es la acertada. Pero podemos darnos cuenta de que por mucho tiempo que le dedicasen al asunto lo cierto es que las dos respuestas tienen el mismo valor de verdad para ellos: nulo. Aunque diesen con la respuesta correcta y finalmente todos estuviesen convencidos de que el cielo es de color azul, lo cierto es que seguirían sin saber cuál es el color del cielo, porque «azul» es sólo un nombre para el efecto subjetivo que los rayos de luz de determinada longitud de onda tienen sobre el sistema visual de los seres humanos. Dado que los alienígenas de nuestro experimento carecen de un sistema visual equivalente todas sus afirmaciones sobre el color son vacías. Del mismo modo, dado que nosotros carecemos del punto de vista de Dios, cualquier afirmación sobre la naturaleza última de la REALIDAD tiene un valor de verdad inexistente. Si nos obstinamos en mantener una teoría de la verdad como correspondencia, entonces ninguna de nuestras teorías científicas es verdadera ni falsa. Sencillamente carece de sentido aplicarles dicho criterio.

El problema que surge al renunciar a la actitud realista en favor de una actitud pragmática es que la utilidad no es un valor absoluto, sino que depende de nuestras intenciones. Ante la afirmación de que algo es útil siempre es pertinente la pregunta «¿útil para qué?». De este modo podemos darnos cuenta de que unas explicaciones son útiles para construir aviones mientras que otras son útiles para fomentar el respeto a los

derechos humanos. Unas son útiles para fabricar bombas y otras son útiles para salvar vidas. Unas sirven para fomentar la tolerancia y otras para estimular el odio hacia los que son diferentes. Y en principio no hay motivos para pensar que tenga que haber una única explicación útil para todos los aspectos de la vida humana. Aún así algunos realistas insisten en que todas las explicaciones válidas –ya veremos más adelante cómo podemos entender «válidas»– tienden a coincidir en líneas generales, por lo que podría ser que en última instancia todas ellas convergieran en una única teoría unificada que finalmente se correspondería con la REALIDAD.

Contra esta postura podemos señalar que pese a la interpretación que solemos hacer de otras teorías, lo cierto es que si dos explicaciones son realmente diferentes –si una no es una simple extensión de la otra– entonces son inconmensurables, esto es, incomparables, y por tanto no pueden converger en ningún punto. Dado que una explicación o el mundo postulado por la misma no son independientes del lenguaje utilizado para expresarlos debemos tener en cuenta que los términos de un lenguaje adquieren su significado de forma holística según las relaciones que guardan con el resto de términos. La palabra «árbol» significa «árbol» no porque se refiera a una entidad externa presente en la REALIDAD, sino porque guarda las relaciones apropiadas con otros términos del sistema como «madera», «hojas», «vegetal», «bosque», etc. Del mismo modo, tomando un ejemplo de la película *Los dioses deben estar locos* (Jamie Uys, dir., 1980), un objeto sólo es una botella de Coca-Cola si podemos relacionarlo con otros recipientes que contienen líquidos y que denominamos botellas y con una serie de empresas que en una economía capitalista venden distintos tipos de bebida bajo nombres o marcas comerciales. Si por el contrario no somos capaces de establecer dichas relaciones porque pertenecemos a una tribu nómada del desierto del Kalahari que no ha tenido contacto con la civilización occidental entonces el mismo objeto que para otros es una botella de Coca-Cola para nosotros puede ser un regalo envenenado de los dioses que si bien puede servir para muchas cosas –para hacer fuego, para machacar bayas, para soplar y hacer música– también puede ser el origen de envidias y luchas que dividan a la tribu. Pero dado entonces que los objetos no son independientes del resto de entidades con las que se relacionan en un determinado mundo, y que adquieren por tanto su existencia no por su supuesto referente REAL sino gracias a los términos y creencias que llevan asociados, entonces no podemos decir que un mismo hecho o un problema exista en dos mundos diferentes y que en última instancia se pueda llegar a un acuerdo sobre los hechos que existen y sus descripciones apropiadas, pues cada mundo

define sus propios hechos.⁴ Volviendo otra vez sobre las teorías astronómicas, aunque exista un objeto llamado Tierra tanto en el mundo postulado por el modelo ptolemaico como en el postulado por el modelo copernicano, lo cierto es que una es irreducible a la otra. La Tierra copernicana se mueve, mientras que la ptolemaica permanece quieta, y pese a lo que nos pueda parecer desde nuestra perspectiva, la presencia o ausencia de movimiento no es trivial en este caso, sino que se trata de una característica esencial en los dos mundos que nos ocupan. De este modo para alguien convencido de la verdad del modelo geocéntrico la afirmación de que la Tierra se mueve no es simplemente falsa, sino que es un sinsentido (Kuhn, 1962, pp. 232-233). «Tierra» significa dos cosas distintas para un habitante del mundo ptolemaico y para un habitante del mundo copernicano, por lo que aunque parezca que los dos están de acuerdo en que existe un objeto llamado Tierra que difiere en sus propiedades, lo cierto es que hablan de dos Tierras distintas.⁵

Aún podría argumentarse que el modelo ptolemaico sencillamente es incorrecto, por lo que no importa si es irreducible o no al copernicano. Sin embargo, aunque esto fuese cierto, parece evidente que existen descripciones válidas del mundo que son incompatibles entre sí. Así por ejemplo si describimos el mundo desde una postura eliminativa diremos que una persona experimenta la excitación de sus fibras C, mientras que si lo describimos desde un marco de corte más fenomenológico diremos que experimenta dolor. Ambas descripciones son válidas o verdaderas, pero sin embargo son incompatibles entre sí, pues el eliminativismo implica la no-

4 A este respecto puede recordarse lo dicho sobre la relatividad ontológica defendida por Quine en el primer capítulo, donde señalaba que la idea de referencia carece de sentido excepto cuando se aplica a un sistema de referencias previo. Del mismo modo la inconmensurabilidad de los paradigmas kuhnianos que estoy siguiendo se puede relacionar con la tesis de la traducción radical de Quine, autor quizá más leído por los filósofos analíticos de la mente.

5 Véase el siguiente fragmento de la novela *El difunto Matías Pascal* de Luigi Pirandello:

—[...] ¡Maldito sea Copérnico!

—¿Pero a ver, qué pinta aquí Copernico? —exclama el padre Eligio, subiéndose la falda del hábito, el rostro encendido bajo el maltrecho sombrero de paja.

—Pinta, padre, pinta. Porque cuando la tierra no giraba...

—¡Y dale! ¡Pero si siempre ha girado!

—No es verdad. El hombre no lo sabía, y por lo tanto era como si no girase. Para mucha gente, aún hoy, no gira. Se lo hice saber el otro día a un viejo campesino, y ¿sabe qué me contestó? Que era una buena excusa para los borrachos. (Pirandello, 1904, pp. 71-72).

existencia de la dimensión subjetiva de la experiencia. Por tanto debemos admitir que si las dos descripciones son verdaderas entonces deben referirse a mundos distintos, pues dos verdades contradictorias no pueden coexistir en una misma realidad. Si existen distintas descripciones válidas del mundo, entonces existen múltiples mundos reales (Goodman, 1989, p. 83). Se hace evidente así el motivo por el que no es posible efectuar una reducción de una perspectiva fenomenológica a una materialista, ya que las realidades postuladas por ambos tipos de discurso son diferentes e incommensurables.

Si lo que estoy exponiendo es cierto y ninguna descripción del mundo puede proponerse como *la única y verdadera*, entonces los intentos de los filósofos de la mente por encontrar una explicación materialista de la conciencia que veíamos en el capítulo anterior carecen de sentido. Si todo el pensamiento filosófico occidental se ha construido sobre la oposición entre lo físico y lo mental es imposible hacer ahora que lo segundo se reduzca a lo primero, pues por definición son dos categorías irreconciliables. Sin embargo esto no significa que las propiedades mentales no existan, como defienden los partidarios del materialismo eliminativo, pues la explicación física de la realidad no es la única sino sólo una de entre tantas posibles, y debe por tanto convivir con otras igual de válidas. En cualquier caso esto tampoco impide que en el futuro pueda desarrollarse una explicación única de todos los aspectos de la existencia humana –lo que sí impide es que dicha explicación se pueda llegar a confundir con la REALIDAD–, pero ésta deberá conllevar necesariamente un cambio radical de nuestra forma de ver el mundo, incluyendo la idea de que las piedras son el paradigma del material del que está construido el universo y que las entidades mentales tienen un carácter «fantasmático». Hasta entonces no tenemos ningún motivo para intentar reducir las entidades fenoménicas o sociales a una explicación física y en consecuencia el problema de la identidad personal podría abordarse desde la presuposición de que nuestros conceptos de «ser humano», «Yo» y «persona» se refieren a nosotros mismos bajo distintas descripciones del mundo, motivo por el cual nuestras intuiciones al respecto parecen incompatibles dentro de un marco materialista. Pero volveré sobre esto más adelante.

4.3. El juego de la vida y la actitud interpretativa

Calificar mi posición de «anti-realista» mientras sostengo que existe una REALIDAD objetiva puede resultar confuso, por lo que quizá resulte

esclarecedor poner mi postura en relación con la de los dos autores que, habiendo teorizado sobre la identidad narrativa, más cerca se encuentran de mis presupuestos: Jerome Bruner y Daniel Dennett.

Como psicólogo, Bruner no se adentra en razonamientos filosóficos profundos. Pero no deja de hacer referencia a sus bases ontológicas, que no son otras que las del constructivismo de Goodman (Bruner, 1987, pp. 11-12), autor que también influye fuertemente en mi propia postura. Para ambos autores la realidad no es más que las descripciones que nosotros mismos hacemos de ella. Haciendo referencia a la novela de Italo Calvino *El caballero inexistente*, del mismo modo que el caballero Agilulfo no existe detrás de su armadura, tampoco nosotros existiríamos detrás de nuestras palabras, por lo que al describirnos a nosotros mismos en tanto Yo es como productos de una narración estaríamos creando aquello que somos. Aunque nuestra narrativa no sea más que una construcción cultural, en última instancia nos convertimos en el producto de nuestra historia (Bruner, 1987, p. 15), ya que no existe ningún referente REAL que se oculte tras ella. Dado que no existe un mundo REAL detrás de nuestras palabras, éstas se convierten en la realidad. Y dado que nuestras palabras pueden construir de un modo válido descripciones incompatibles del mundo –otra vez el dolor y la excitación de las fibras C–, debemos aceptar la existencia de muchos mundos en los que existen –y existimos– objetos de distinta naturaleza.⁶

Por su parte la postura de Dennett es algo más complicada, pues confiere un estatus ontológico distinto a las entidades físicas y a las mentales, con respecto a las que él mismo se cataloga de «realista débil» o «semi-realista» (Dennett, 1991b, pp. 27, 30). Esto es así porque las entidades físicas para Dennett existen en el nivel fundamental de la realidad, por lo que al describir el mundo en términos físicos agotamos su existencia. Sin embargo cuando tratamos con entidades complejas como podemos ser los seres humanos una descripción física de nuestro comportamiento resultaría impracticable, por lo que debemos recurrir a explicaciones que, pasando por encima del nivel físico, detecten patrones observables a un nivel más elevado (Dennett, 1991b, p. 41). De este modo, aunque el comportamiento humano podría ser explicado por los cambios producidos en el nivel neuronal, es más sencillo explicarlo tomando una actitud interpretativa diferente: la actitud intencional. Dennett explica de la siguiente forma el modo en que se debe proceder para tomar esta actitud:

[...] primero se decide tratar al objeto cuyo funcionamiento hay que predecir como un agente racional; luego se deduce qué creencias debería tener ese agente, dada su posición en el mundo y su objetivo. Más tarde se deduce qué

6 Para una discusión sobre el constructivismo de Goodman puede verse el artículo de Xavier de Donato-Rodríguez (2009).

deseos tendría que tener siguiendo las mismas consideraciones, y por fin se predice que este agente racional actuará para conseguir sus metas a la luz de sus creencias. (Dennett, 1987, pp. 28-29)

Mediante esta nueva forma de acercarse a nuestra vida la actitud intencional detecta patrones de comportamiento que, aunque no poseen una fiabilidad del cien por cien y no pueden ser reducidos exactamente a una explicación física debido a la indeterminación provocada por la tesis de la traducción radical de Quine (Dennett, 1991b, pp. 46-48), son reales en cierto sentido, pues son útiles para predecir y explicar los cambios que se producen en la realidad. De este modo Dennett concluye que su postura se encuentra a medio camino entre el realismo y el instrumentalismo (Dennett, 1991b, p. 51), lo que explica la diferenciación que hace entre los Yoes y los personajes literarios en cuanto a su carácter ficticio que mencionábamos en la nota al pie 26 del capítulo dos: mientras que los personajes literarios no existen en absoluto y son por tanto plenamente ficcionales, los Yoes son una entidad existente cuando adoptamos la actitud intencional para interpretar el mundo, por lo que aunque no son tan reales como los átomos, sí que al menos son semi-reales (*vid.* Muñoz Corcuera, 2013). Dennett explica su postura recurriendo a una analogía con el juego de la vida.

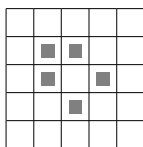
El juego de la vida es un autómatas celular⁷ diseñado por el matemático John Horton Conway en 1970. Consta de un tablero virtualmente infinito dividido en cuadrículas o células, de modo que cada célula está rodeada por ocho células vecinas. Las células pueden estar en dos estados: encendidas o apagadas. La única aportación del jugador al juego consiste en elegir qué células estarán encendidas o «vivas» al comienzo de la partida y cuáles estarán apagadas o «muertas», pues una vez se inicia el juego la disposición de las cuadrículas en el tablero evoluciona automáticamente por turnos siguiendo tres sencillas reglas:

- Si una célula muerta tiene a su alrededor exactamente tres células vivas, en el siguiente turno «nace» (en el turno siguiente estará viva).
- Si una célula viva tiene a su alrededor exactamente dos o tres células vivas, en el siguiente turno sigue viva.
- En todos los demás casos, en el siguiente turno las células vivas mueren y las muertas permanecen muertas.

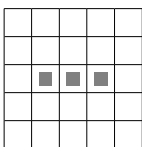
⁷ Un autómatas celular es un modelo matemático para un sistema dinámico que evoluciona en pasos discretos. En cualquier caso la explicación del juego es más clara que su definición formal.

Gracias a la invariabilidad de las reglas, dado un estado inicial cualquiera una persona puede saber cuál será el estado del tablero en el siguiente turno sin ningún tipo de duda. La evolución *física* del sistema puede ser conocida del mismo modo a como el demonio de Laplace podría conocer la evolución del universo, pues al ser un sistema causalmente cerrado que evoluciona siguiendo unas reglas simples no cabe duda de que su destino está determinado. Sin embargo, pese a su aparente simplicidad, el juego adquiere rápidamente dimensiones que lo complican, como veremos a continuación.

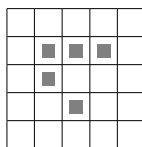
Un primer aspecto que llama la atención en el juego de la vida es que algunas configuraciones iniciales tienen patrones de evolución más interesantes que otras. Así por ejemplo mientras algunas configuraciones evolucionan caóticamente o desaparecen en pocos turnos, las de la figura siguiente, pese a su simplicidad, tienen comportamientos llamativos: la configuración llamada «barco» permanece estable; «parpadeador» oscila entre formar una línea horizontal y una vertical; y «planeador» se mueve en diagonal hacia arriba y hacia la izquierda.



Barco



Parpadeador

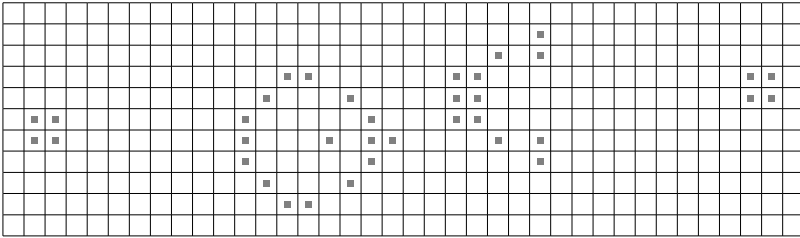


Planeador

Pero no sólo estas figuras simples tienen comportamientos llamativos, sino que existen infinidad de objetos ya conocidos que tienen comportamientos igualmente predecibles: hay otros «osciladores» además de los parpadeadores, como pueden ser los sapos; «deslizadores» que se mueven del mismo modo que los planeadores, como las naves espaciales; «pistolas», que son estructuras fijas que a su vez generan planeadores o naves; «locomotoras», que se mueven como los deslizadores pero dejando a su paso un rastro de humo; «comedores» que son capaces de hacer desaparecer otros objetos que entran en colisión con ellos, etc.

Lo interesante de estas configuraciones es que al empezar a jugar casi sin darnos cuenta se produce un deslizamiento ontológico en nuestra forma de describir el juego, pues al hablar de objetos que permanecen estables, que parpadean o que se mueven dejamos de referirnos al nivel físico básico para hacerlo a un nivel de *diseño* más elevado. Este cambio simplifica nuestra comprensión de la evolución de una determinada configuración inicial, pues nos evita tener que calcular «manualmente» lo

que sucede en el nivel físico, lo cual puede ser de gran ayuda cuando nos enfrentamos no a figuras simples aisladas, sino a enormes conjuntos de ellas. Sin embargo esta simplificación tiene un precio, pues la fiabilidad de nuestras predicciones disminuye con respecto al nivel físico según aumenta la complejidad de los sistemas. Por poner un ejemplo sencillo, aunque sabemos que un comedor es capaz de eliminar un planeador sin ningún problema, si ponemos una pistola de planeadores como la de la figura siguiente enfrente de un comedor no sabemos si éste tendrá tiempo de eliminar todos los planeadores o si se irá produciendo un desfase que finalmente acabará provocando el colapso del comedor o incluso de la propia pistola.



La pistola de planeadores de Gospel es la pistola más sencilla del juego. Los dos objetos centrales se mueven de izquierda a derecha «rebotando» en los bloques exteriores de tal modo que cada vez que los dos objetos centrales chocan entre sí se genera entre ellos un planeador que se mueve en diagonal hacia abajo y hacia la derecha.

A pesar de este cambio ontológico que se produce en nuestra forma de hablar al abandonar el nivel físico, podemos preguntarnos si realmente existen los objetos de los que hablamos en el nivel de diseño. Cuando creamos un planeador, ¿existe realmente un objeto que se mueve? ¿O únicamente existe un patrón de cambio en los estados de las células? Y en el caso de que sólo exista un patrón, ¿se puede decir que el patrón es real? ¿O sólo existe en nuestra mirada? El problema se complica aún más cuando conocemos otra de las características más llamativas del juego: en su tablero se puede construir una máquina universal de Turing.

Una máquina de Turing es un dispositivo capaz de manipular símbolos en base a una serie de reglas de tal modo que, pese a su simplicidad, se puede configurar para simular la lógica de cualquier algoritmo computable. Es decir, una máquina de Turing puede computar cualquier proceso que pueda ser expresado como un conjunto finito de instrucciones ordenadas y bien definidas, como por ejemplo cualquier programa de

ordenador. De este modo, en el juego de la vida podemos construir una máquina de Turing capaz de ejecutar cualquier *software*, como por ejemplo un juego de ajedrez. Esta máquina, cuya construcción requeriría miles de millones de células, sería capaz de mantener una partida de ajedrez consigo misma o contra un jugador humano, de modo que cada vez que se le suministrase la información sobre la posición de las fichas en el tablero devolviese información sobre qué movimiento desea hacer. Es importante notar que dado que no tendríamos un interfaz gráfico que interpretase visualmente en clave ajedrecística los movimientos de la máquina, ante nuestros ojos únicamente veríamos millones de células encendiéndose y apagándose siguiendo un patrón físico desconocido y enormemente complejo. Sin embargo, pese a que tanto el nivel físico como el de diseño presentarían patrones casi indescifrables para nuestras limitadas capacidades, podríamos dar un salto a un nuevo nivel interpretativo que nos permitiese suponer cuáles serían los movimientos de ajedrez que devolvería la máquina sin necesidad de prestar atención a los patrones de niveles inferiores: podríamos adoptar la misma actitud *intencional* que adoptaríamos al enfrentarnos a un jugador humano, y suponer que la máquina es un ser racional que desea ganar la partida y por tanto realizará los mejores movimientos disponibles para lograr un jaque mate. Por supuesto podríamos equivocarnos en nuestras suposiciones, del mismo modo a como podemos equivocarnos al prever los movimientos de cualquier jugador humano, pero sin duda al adoptar esta postura interpretativa tendríamos un nivel aceptable de aciertos, por lo que podemos decir que si nos situamos en el nivel intencional podemos detectar patrones de conducta en la máquina que no son visuales, sino psicológicos o intelectuales (Dennett, 1991b, pp. 40-42). La pregunta vuelve a hacer aparición con más fuerza que antes: ¿existen realmente los patrones? ¿O somos nosotros los que nos los inventamos?

Completando la analogía entre el juego de la vida y el mundo real, podemos identificar el nivel físico del juego con el nivel subatómico de la materia, el nivel de diseño con el de los objetos macroscópicos y el intencional con el de la psicología humana. En este momento alguien podría pensar que los patrones existen siempre y cuando se correspondan con algo real, de modo que la máquina de Turing que juega al ajedrez no tiene en realidad creencias ni deseos, ya que no tiene una mente –no tiene representaciones internas– pero que los seres humanos sí las tienen. Sin embargo para Dennett sería un error pensar que la justificación para describir algo desde la actitud intencional depende de la existencia de alguna característica oculta que nos permite aplicársela a los seres humanos pero no, por ejemplo, a objetos inanimados. Según la teoría de

Dennett podemos encontrar patrones tanto en los niveles físico como de diseño e intencional en el comportamiento de prácticamente cualquier entidad, y por tanto estamos justificados para aplicar la actitud intencional siempre que obtengamos alguna capacidad de predicción que no estuviese a nuestro alcance en otro nivel. Así por ejemplo aunque podríamos aplicar la actitud intencional a una estatua, presuponiendo que dado que desea ser el centro de atención y ser observada permanecerá quieta, dicha interpretación no nos da acceso a un poder predictivo que no tuviésemos en el nivel de diseño o en el físico, por lo que es preferible no utilizarla (Dennett, 1987, p. 33). Sin embargo si hablásemos de un ordenador, o incluso de un termómetro un poco complejo, sí que podría ser útil asignarle creencias y deseos para predecir su comportamiento. No se trata por tanto de que la atribución de cualidades mentales dependa de la existencia de representaciones internas, sino que cuando la estrategia intencional funciona para un objeto somos nosotros los que nos esforzamos en interpretar alguno de sus estados o procesos internos como representaciones (Dennett, 1987, p. 41).

¿Los patrones son entonces algo real o una ficción útil que nos inventamos para comprender las cosas que no entendemos? La postura de Dennett al respecto es ambivalente. Por un lado defiende que los patrones son reales y que además no dependen de la mirada de un observador para existir, sino que lo hacen incluso aunque sean invisibles para nosotros. Por otro, no dejan de ser una ficción útil –como los centros de gravedad– y por tanto no son tan reales como los objetos físicos. La idea básica es que en una serie de datos existe un patrón por cada forma posible de «comprimir» la información necesaria para transmitir esos datos (Dennett, 1991b, p. 34). De este modo en toda situación existen objetivamente infinidad de patrones que somos incapaces de detectar o que no consideramos relevantes –nuestra evolución genética y cultural ha determinado aquellos patrones que nos son accesibles, como por ejemplo los patrones intencionales– pero que sin embargo podrían estar disponibles para otras especies. En este sentido es importante señalar que la existencia de un patrón en una serie de datos no excluye la existencia objetiva de otros patrones, independientemente del grado de imperfección que presente cualquiera de ellos (Dennett, 1987, p. 38). No hay unos patrones más «verdaderos» que otros aunque unos tengan mayor capacidad predictiva, sino que todos tienen el mismo grado de existencia semi-objetiva.⁸ Esto es

8 Sin embargo esto no es un relativismo en el que cualquier postura interpretativa sea válida –Dennett pone especial interés en distanciarse de posturas perspectivistas radicales como la de Rorty (Dennett, 1987, p. 46)– ya que la existencia de patrones se puede comprobar objetivamente mediante la

así porque si dos patrones funcionan, es decir, si tienen un alto grado de fiabilidad en sus predicciones, el hecho de que uno admita la existencia de un menor porcentaje de ruido que el otro –que uno vea regularidades donde otro sólo ve interferencias aleatorias y por tanto tenga mayor cantidad de aciertos– no implica que uno sea más verdadero u objetivo, pues debido a la indeterminación que implica la tesis de la traducción radical de Quine es posible que los aciertos de uno no puedan servir para modificar o mejorar el otro y viceversa. Un patrón acertará en unos casos, otro acertará en otros, en algunos acertarán los dos y en otros no acertará ninguno, y puede que no haya posibilidad de encontrar un patrón unificado que acierte en todas las ocasiones (Dennett, 1991b, p. 48).

El problema para Dennett es cómo compaginar esta actitud pragmática con respecto a los patrones con su afirmación de que existen realmente, pues la opción semi-realista que plantea parece incoherente. Es decir, al otorgar una existencia más real al nivel físico, la relación de éste con los otros niveles semi-reales se vuelve oscura. Así, mientras que mantiene que en última instancia alguna variante del funcionalismo será cierta, pues se podrán encontrar relaciones entre el nivel neuronal y el intencional que permitirían efectuar una reducción casi perfecta del uno al otro (Dennett, 1987, p. 43), también asegura que existen verdades que no pueden ser captadas a no ser desde el nivel intencional, lo que parece implicar que por tanto son irreducibles (Dennett, 1987, pp. 35-36). Esta inconsistencia se traduce en que las entidades mentales, como los Yoes, son para Dennett reales o ficticias según qué le conviene más en cada momento, dando la sensación de que se trata de una postura incoherente.

Una manera de solucionar este problema de la ontología de Dennett sería transformar su realismo débil en uno aún más débil, en el que la frontera con el anti-realismo de Goodman esté más cerca. Si aceptamos las críticas de Goodman y las que he venido presentando en las últimas páginas a la posibilidad de llegar a conocer la REALIDAD, entonces la tentación de Dennett de jerarquizar los niveles de interpretación y otorgar un estatus más REAL a las entidades físicas que a las mentales desaparece. No se trata de defender que la REALIDAD no exista, tal y como sostiene Goodman, sino de darse cuenta de que sea lo que sea la REALIDAD,

posibilidad o imposibilidad de comprimir la información. En este sentido Dennett señala que la actitud intencional defendida por él tiene una alta capacidad predictiva, mientras que la actitud astrológica –el intento de predecir el comportamiento de las personas según la posición de los astros el día de su nacimiento– es incapaz de predecir nada a no ser por casualidad, pese a que para Rorty la actitud intencional y la astrológica tendrían el mismo valor (Dennett, 1991b, p. 50).

nosotros somos incapaces de percibirla tal y como es, de modo que lo que describimos desde la actitud física no es la REALIDAD en sí, sino patrones físicos de comportamiento en lo que quiera que ésta sea.⁹

Si aceptamos esto entonces las entidades físicas no tienen un estatus privilegiado con respecto a las mentales, pues ambas dependen de la existencia de patrones observables por nosotros, al mismo tiempo que la actitud instrumentalista, pragmatista o constructivista que hemos adoptado no desemboca en un relativismo radical –postura que como dijimos normalmente se asocia con el anti-realismo–, pues las entidades que podemos considerar reales según las distintas actitudes interpretativas, teniendo en cuenta que su postulación debe permitirnos hacer predicciones útiles y fiables, están restringidas por lo que quiera que en última instancia sea la REALIDAD.

4.4. Tipos de mundo

Antes de explorar cómo el problema de la identidad personal puede encontrar su solución en el marco que estoy exponiendo será útil ahondar un poco más en los distintos tipos de mundo que podemos construir.¹⁰

9 Una actitud pragmatista parecida a la de Dennett es la del psicólogo evolutivo Steven Pinker, para quien «un ser humano es al mismo tiempo una máquina y un agente libre capaz de sentir, dependiendo de cuál sea el propósito de lo que se estudia, del mismo modo en que es un ciudadano que paga impuestos, un corredor de seguros, un paciente de una clínica dental o doscientas libras de lastre en un avión, dependiendo siempre de los propósitos que guían un estudio. La postura mecanicista nos permite comprender qué nos hace funcionar y cómo nos ajustamos al universo físico. Pero fuera del ámbito de este estudio, hablamos unos con otros como seres humanos dignos y libres» (Pinker, 1997, p. 83). Por otro lado, dadas las fervientes críticas que hace Pinker contra las corrientes filosóficas postmodernas que defienden que el mundo es construido por los seres humanos, y su defensa de que nuestras categorías mentales se adaptan al mundo natural porque la mente evolucionó para que se adaptasen, es de suponer que también se sentiría atraído por la teoría de la detección de patrones genética y culturalmente determinada de Dennett (Pinker, 1997, pp. 84, 399-400). En cualquier caso, y a diferencia de Dennett, Pinker admite que no existe una explicación científica de la conciencia o de las cualidades morales, cuestión que reconoce misteriosa y más adecuada para otros tipos de investigación (Pinker, 1997, p. 196).

10 Con la expresión «construir un mundo» me refiero al hecho de que al realizar una descripción postulamos la existencia de un mundo externo a nosotros al cual nuestra descripción se refiere. En este sentido los mundos en los que vivimos no existen hasta que postulamos su existencia al describirlos, aunque

Especialmente porque como hemos dicho el anti-realismo se presta a interpretaciones del tipo «todo vale», y aunque ya hemos avanzado bastante en la dirección de demostrar que esto no es así, la resistencia a la que frecuentemente se enfrentan los anti-realistas aconseja profundizar un poco más. Para ello trazaré tres divisiones. La primera es la más importante y diferenciará entre mundos válidos e inválidos. La segunda, que hasta cierto punto correrá pareja a la primera, distinguirá entre mundos científicos y artísticos, cuestión que tendrá cierta importancia para decidir si es apropiado utilizar como modelo a personajes literarios para entender a las personas. La última nos permitirá separar la realidad y la ficción.

Ya hemos señalado anteriormente que para que un mundo –o una interpretación de la REALIDAD– sea válido únicamente es necesario (1) que tenga utilidad para nuestros propósitos. Especialmente en el sentido de que dé lugar a explicaciones y predicciones fiables sobre su comportamiento. Siguiendo a Dennett además podemos decir que las predicciones serán fiables cuando se basen en la existencia de patrones de algún tipo, de modo que las interpretaciones arbitrarias y de nulo poder predictivo o explicativo, como por ejemplo las astrológicas, no darán lugar a mundos válidos. La utilidad es el único requisito realmente imprescindible para la validez de un mundo, y por tanto cualquier explicación útil dará lugar a un mundo válido, mientras que por el contrario cualquier explicación inútil dará lugar a un mundo inválido. Sin embargo, para ayudarnos en la exposición y en la diferenciación entre mundos científicos y artísticos, a este requisito podemos añadir además que para ser válido un mundo tiene que (2) ser verosímil, (3) tener coherencia y lógica internas, (4) estar abierto a nuevos hechos y por último (5) debe incluirnos a nosotros en su seno. En el fondo estos cuatro nuevos requisitos están relacionados con (1), pues difícilmente daría lugar a predicciones fiables un mundo completamente inverosímil e incoherente en el que no se aceptase la cabida de nuevos hechos y en el que además nosotros no existiésemos –éste sería quizá el mundo postulado por una novela fantástica de pésima calidad–. Sin embargo, como acabamos de decir, en vistas a distinguir entre mundos científicos y artísticos nos puede ser de ayuda destacar estos aspectos menores que contribuyen a la utilidad

Algo importante que debemos destacar con respecto a los mundos válidos y las cinco características que acabamos de señalar es que existen ciertos límites a la aplicación de (2) y (3), pues la verosimilitud de algo

en un sentido literal lo que nosotros construimos no es un mundo, sino la descripción del mismo. Véanse a este respecto las críticas de Scheffler al constructivismo de Goodman (Scheffler, 1980).

depende en gran medida de la percepción subjetiva del intérprete —alguien puede considerar inverosímiles las consecuencias de la mecánica cuántica y pensar que nuestra física debe estar equivocada, mientras que otro las puede aceptar como creíbles porque el razonamiento lógico-matemático que las sustenta le parece suficiente prueba—, mientras que en todo mundo existe un grado tolerable de incoherencias. Sobre el problema de la verosimilitud volveré más adelante, por lo que no diré nada más por el momento. En cuanto al de la coherencia y la lógica internas es pertinente indicar que un mundo no tiene por qué tener una explicación de todo para ser válido —recordemos que todos los patrones tienen que admitir la existencia de un porcentaje de ruido, es decir, de datos inexplicables según el propio patrón—, por lo que puede aceptar en su seno algunas inconsistencias que no se consideren lo suficientemente importantes como para motivar su abandono o que se piense que podrían ser resueltas si se investigasen adecuadamente. En cualquier caso siempre es posible que en un determinado momento dichas incoherencias lleguen a ser insoportables a la luz de nuevos descubrimientos, motivando que lo que antes era considerado un mundo válido con un porcentaje de incoherencias aceptable pase a ser inválido al dejar de ser coherente en el grado necesario.

Dejando a un lado estas matizaciones, ya hemos dicho que la distinción entre mundos válidos e inválidos corre hasta cierto punto pareja a la distinción entre mundos científicos y artísticos, pues los mundos científicos son válidos —recuerdo que estoy entendiendo «ciencia» como sinónimo de «conocimiento»— y en general los mundos artísticos son inválidos. Sin embargo no todos los mundos inválidos son artísticos —por poner un ejemplo, la pésima novela de la que hablaba hace unos párrafos seguramente no tiene ningún valor artístico—. A este respecto trataré de indicar algunas características que tienen los mundos artísticos que, pese a no ser suficientes para hacerlos calificar como válidos, los acercan a éstos. En cualquier caso no podemos olvidar que lo más característico del arte es precisamente su resistencia a la definición y su tendencia a la transgresión de los límites. El arte cambia constantemente, y en consecuencia su definición se agota en explicitar que arte es aquello que se considera arte. En este sentido la caracterización que esbozaré al respecto puede interpretarse como una guía para interpretar qué es lo característico del arte clásico, con algunos apuntes sobre los motivos que impiden mantener dichas características cuando se consideran obras (post)modernas.

En primer lugar, frente a lo que pudiera pensarse en un primer momento, (2) y (3) son requisitos que en general cumplen los mundos

artísticos. En una película, pongamos por caso, pueden suceder hechos fantásticos que desafíen la lógica de nuestro mundo, como que un hombre pueda volar. Sin embargo, para que dicha obra sea considerada creíble y no una simple acumulación de hechos sin sentido, el vuelo tiene que estar justificado internamente de alguna forma –el hombre que vuela puede hacerlo porque no es un ser humano, sino un extraterrestre del planeta Krypton–. La lógica interna del relato –que no la lógica de algún otro mundo externo– tiene que permitir que se den esa clase de sucesos y explicar cómo son posibles para que el lector no se sienta engañado, como sucedería por ejemplo si Clark Kent fuese en verdad simplemente el hijo de una pareja de granjeros humanos y sus poderes sobrenaturales no tuviesen ninguna explicación. En este sentido cualquier suceso que aparezca en una película tiene que ser coherente con el resto del texto. Si las primeras escenas establecen como una verdad incontrovertible que Superman es Clark Kent, no sería aceptable que más adelante dicha verdad se contradijera. Por otro lado, aunque en principio la lógica interna del relato pueda ser todo lo diferente que se quiera de la lógica que impera en nuestro mundo, existe una delicada línea que separa aquello que es verosímil de aquello que no y ésta no siempre tiene que ver con la presencia de elementos fantásticos. Por poner un ejemplo, los escritores noveles a menudo se interesan por historias reales que les resultan llamativas por la asombrosa concatenación de casualidades que le sucedieron a una determinada persona. Sin embargo cuando tratan de convertir dicha historia en una novela se encuentran con que el lector no la acepta como verosímil. Aunque realmente sucediera, la acumulación de sucesos asombrosos en una novela puede acabar pareciendo una tomadura de pelo. En cualquier caso, como señalé anteriormente, volveré sobre el problema de la verosimilitud en el sexto capítulo. De momento lo único que nos interesa es que aunque la verosimilitud y la coherencia y la lógica internas sean condiciones necesarias para la validez de un mundo, no son condiciones suficientes, pues son dos de las características más destacadas de los mundos artísticos.

Por su parte las tres condiciones restantes para la validez de un mundo sí que pueden sernos de más ayuda para diferenciar los mundos artísticos de los mundos válidos. Como hemos venido diciendo, (1) es la característica más sobresaliente de los últimos. La utilidad es en gran medida la que motiva nuestra aceptación o rechazo de un determinado mundo, pues es preferible una mentira útil que una verdad irrelevante (Goodman, 1978, pp. 164-165). Por el contrario las obras de arte, aunque pudieran presentar utilidad para algún propósito, rara vez éste es el fin que

persiguen, pues su interés no suele dirigirse hacia objetivos prácticos sino hacia (6) el goce estético.¹¹

En estrecha relación con la utilidad se encuentra (4). Mientras un mundo válido necesita estar abierto para poder dar explicación a los nuevos hechos que se producen todos los días en nuestra vida, los mundos artísticos generalmente se nos presentan cerrados, como si el mundo hubiese terminado y no quedasen más hechos por explicar (*cf.* Dennett, 1988, p. 109). En este sentido cualquier mundo que consideremos válido tiene que ser capaz de explicar lo que nos sucederá dentro de cinco minutos, mientras que el mundo de Sherlock Holmes se nos presenta clausurado y la pregunta por el mañana no tiene sentido. Su mundo está construido por un corpus cerrado de textos a los que no es posible añadir nada. En cualquier caso debemos señalar que existen obras artísticas que se presentan abiertas a nuevas aportaciones, especialmente gracias al uso de internet. Un ejemplo podrían ser las wiki-novelas, en las que cualquier lector puede modificar lo que ya está escrito, o añadir nuevas páginas a la historia. Sin embargo, a pesar de la vinculación de (4) con (1) y de que no todos los mundos artísticos se encuentran cerrados, considero que se trata de un criterio que merece la pena ser señalado. El motivo es que además de ser cierto que la gran mayoría de los mundos artísticos están cerrados, para la minoría restante normalmente el hecho de presentarse abiertos se trata de una transgresión explícita de las normas clásicas.

Finalmente (5), el hecho de que nosotros existamos en los mundos válidos, es otra característica fuertemente vinculada a (1), pues un mundo en el que no estuviésemos tendría un escaso poder explicativo para nosotros. En contraste, los mundos artísticos generalmente postulan realidades ajenas en las que, aunque puedan encontrarse referencias a entidades que existen en mundos válidos –una novela puede estar

11 Esto no implica que el arte, y en concreto la literatura, no sirva para nada, o sólo para deleitarnos con su belleza. Tal y como Jorge Volpi señala la literatura tiene o pudo tener un gran valor en nuestra evolución, ampliando las capacidades de nuestro cerebro en tanto mecanismo para predecir el futuro. Las situaciones hipotéticas que presenta la literatura serían como un campo de entrenamiento para la vida real. Al mismo tiempo la literatura sirve como un vehículo de transmisión de las ideas y valores que conforman nuestra idea de lo que es ser lo que somos (Volpi, 2011, pp. 13-32). Por el contrario lo que intento transmitir con la idea de que los mundos del arte no tienen utilidad es que no sirven para predecir las situaciones concretas que van a sucedernos o para conocer lo que sucedió en el pasado, pues no es ese su objetivo. Si queremos saber qué es lo que va a ocurrir mañana ninguna novela nos ofrecerá ninguna pista más fiable que la que nos pueda ofrecer la astrología o un vidente. Si acierta será por casualidad.

ambientada en un New York altamente realista, por ejemplo—, nosotros no existimos. En cualquier caso, al igual que sucede con (4), a pesar de que por norma general nosotros no existimos en los mundos artísticos, existen excepciones que tratan de subvertir las normas clásicas incluyendo a personas reales. Así por ejemplo en el arte postmoderno es habitual mezclar la realidad y la ficción con la intención de mostrar que la distinción entre ambas es arbitraria, pues la realidad no consiste en hechos dados de antemano, sino que es igualmente construida de forma subjetiva. Las novelas de Paul Auster, en las que él mismo aparece de forma frecuente, pueden servir para ilustrar este aspecto.

Haciendo un resumen de lo dicho, los mundos válidos cumplen todos los requisitos que hemos señalado (+1, +2, +3, +4, +5), mientras que cualquier mundo que tenga alguna carencia que afecte a estos aspectos es inválido. Una subclase de los mundos inválidos es la de los mundos artísticos. Éstos, a pesar de ser verosímiles y de tener una coherencia y lógica internas, no tienen una utilidad práctica sino una intención estética, no nos incluyen a nosotros en su seno y se presentan cerrados a nuevos hechos (-1, +2, +3, -4, -5, +6). Sin embargo, como ya hemos señalado, el arte por su propia naturaleza se resiste a la definición y tiende a la transgresión de los límites, por lo que para ser artístico la principal característica que debe cumplir un mundo es (7) ser considerado artístico. Si cumple con esta condición, incluso un mundo válido puede ser artístico.

En cualquier caso lo que más nos va a interesar para nuestra discusión, esto es, para decidir si es legítimo utilizar a personajes literarios como modelos para entender lo que son las personas, es el hecho de que los mundos artísticos cumplan con los requisitos (2) y (3). El resto de requisitos implican que, por norma general, los mundos postulados en las obras literarias van a ser distintos al nuestro, no van a intentar detectar patrones en la REALIDAD sino a inventarse una realidad alternativa, por lo que los hechos concretos que presenten no serán relevantes. Pero sin embargo el cumplimiento de (2) y (3) apunta a que los mundos artísticos van a compartir con los mundos válidos las herramientas y la forma en que se construyen, por lo que los personajes literarios pueden servir como un modelo simplificado de aquello que son las personas. En cualquier caso le dedicaré mayor atención a este punto en el sexto capítulo.

Finalmente, antes de dar por concluido este apartado, la última distinción entre mundos que me gustaría aclarar es la que se puede trazar entre mundos ficticios y reales. En principio podríamos sentirnos inclinados a decir que los mundos válidos son reales y los inválidos ficticios, sin embargo hacerlo podría llevar a confusión, pues dicha distinción parece sugerir que existe algún motivo ontológico de peso a favor de la existencia REAL de los mundos válidos. Pero a pesar de que un

mundo coherente nos pueda parecer más REAL que uno que no lo sea, de hecho no tenemos ningún motivo para creer que la REALIDAD sea coherente o verosímil, por lo que atendiendo al uso que solemos dar en nuestra vida cotidiana a los términos «realidad» y «ficción» creo que es mejor hacer una división distinta.

Como hemos venido diciendo a lo largo de este capítulo, existen distintos mundos válidos incompatibles entre sí, como pueden ser por ejemplo el mundo postulado por un científico actual ateo y el postulado por uno creyente. A pesar de que puedan parecerse en muchos aspectos –podemos imaginar que los dos científicos que estamos imaginando están aparentemente de acuerdo en todo menos en la existencia de Dios–, estos dos mundos son irreconciliables, pues uno contempla la existencia de una entidad que el otro deniega. Si le preguntásemos al científico ateo sobre esta entidad, lo más seguro es que su respuesta fuese algo parecido a «Dios no existe, no es más que una ficción». Ante esta afirmación el científico creyente respondería «Dios sí existe, es real». Esta asociación entre «existir» y «ser real» y «no existir» y «ser ficción» es recurrente en nuestra forma de hablar. Sin embargo, como hemos venido diciendo, no tenemos argumentos para decir que algo existe o deja de existir realmente, sino que únicamente podemos decir si algo existe o no en un determinado mundo subjetivo. De este modo creo que la mejor distinción que se puede hacer entre mundos reales y ficticios no está basada en las características propias de los mundos, sino en nuestra posición con respecto a ellos. De este modo podemos decir que algo es real cuando pertenece al mundo subjetivo desde el que hablamos, mientras que es ficticio cuando pertenece a un mundo diferente. En este sentido parece expresarse David Carr sobre la naturaleza de los elementos ficticios cuando critica que Dennett considere que el Yo tiene un estatus ficcional:

Lo que quiero decir es que el estatus ficcional o no-ficcional del personaje principal de una historia está en función no de si de hecho se corresponde con algo en el «mundo real», sino del sentido del texto según es leído o escuchado y comprendido. Ahora volvamos a esas historias que ustedes y yo contamos sobre nosotros mismos. Estos son los textos que Dennett, en su rol como heterofenomenólogo, quiere leer e interpretar. En mi opinión lo que los hace autobiográficos y no ficcionales es que yo, el narrador, asumo que el personaje principal de estas historias no sólo existe, ¡sino que soy yo! En otras palabras, me identifico con el personaje del que estoy hablando. Las personas ficticias son personas que consideramos que no existen y que, en última instancia, no nos importan. (Carr, 1998, p. 343)¹²

12 La traducción es mía. El original es:

«My point is that the fictional or non-fictional status of the central character in a narrative is a function, not of whether it in fact corresponds

Siendo así, los objetos que existen real o ficcionalmente varían por tanto a lo largo del tiempo y de una persona a otra, pues se encuentran en un equilibrio líquido. Diferentes personajes históricos han vivido en mundos distintos, así como las personas que habitamos un determinado momento histórico podemos vivir en diferentes mundos que compiten entre sí, de modo que lo que para algunos es real para otros es ficción. Lo real y lo ficcional poseen el mismo estatus ontológico objetivo: ninguno. Ambos tipos de entidades existen únicamente desde una perspectiva de primera persona y son por tanto subjetivas. La diferencia terminológica intenta únicamente mostrar el modo en que hablamos en nuestra vida diaria. La realidad y la ficción dependen de nuestra postura, por lo que están siempre sujetas a negociación.

Sin embargo resulta evidente que únicamente los objetos que existen en los mundos válidos pueden ser reales, mientras que los objetos de los mundos inválidos deberán ser siempre ficticios, pues nosotros nunca podremos situarnos plenamente en un mundo inválido. Siendo así creo que podemos establecer una nueva categoría ontológica para referirnos a aquellas entidades que no sólo no son reales en un determinado mundo, sino que no pueden serlo en ninguno. Llamaremos por tanto FICTICIOS a este tipo de objetos, de tal manera que la postura que defiendo sostiene la existencia de cuatro niveles ontológicos distintos:

1. La REALIDAD, que se refiere a aquello que no podemos conocer pero que en verdad es.
2. La realidad, que se refiere a aquello que creemos que existe dentro del mundo subjetivo desde el que estamos hablando.
3. La ficción, que se refiere a aquello que creemos que no existe dentro del mundo subjetivo desde el que hablamos.
4. La FICCIÓN, que se refiere a aquello que racionalmente no podemos creer que exista en ningún mundo válido.

En cualquier caso podemos señalar que no existe ninguna relación entre los objetos REALES y FICTICIOS, pues aunque racionalmente no

to something in the “real world”, but rather of the sense of the text as it is read or heard and understood. Now return to those stories that you and I tell about ourselves. These are the texts that Dennett, in his role as heterophenomenologist, wants to read and interpret. It seems to me that what makes them autobiographical and *not* fictional is that I, the storyteller, take the central character in these stories not only to exist, but to be me! I identify myself, in other words, with the character I am talking about. Fictional persons are persons we take not to exist, and, ultimately, they don't matter to us».

podamos creer en la existencia de estos últimos, al no poder conocer la REALIDAD no podemos saber si ésta es comprensible desde nuestra racionalidad. De este modo podría ser que aunque los unicornios –por poner un ejemplo– no existan en ningún mundo real y por tanto sean FICTICIOS, en última instancia sí que existan REALMENTE.

4.5. El problema de la identidad narrativa

Como señalé al principio de este capítulo no es mi intención analizar hasta sus últimas consecuencias la postura que he tratado de esbozar en las páginas anteriores, por lo que creo que ya he dicho lo suficiente para mis propósitos. El esquema básico ha sido presentado, así como las líneas que deberían seguirse para desarrollarlo en profundidad. Siendo así, creo que es el momento de volver sobre nuestros pasos y recordar que el motivo que me llevó a proponer esta alternativa era la dificultad que encontrábamos en un marco materialista para resolver los problemas a los que se debe enfrentar el NPN. Especialmente la imposibilidad de dar una explicación satisfactoria de cómo podemos ser al mismo tiempo tres entidades reales diferentes –seres humanos, Yoes y personas– que parecen coexistir en un mismo espacio físico pero que son irreducibles entre sí, cuyas condiciones de supervivencia son independientes y con las que podemos identificarnos en el mismo grado de tal modo que nos podemos alegrar en la misma medida por la supervivencia de cualquiera de ellas. Si el nuevo marco resulta propicio para su solución, entonces habremos alcanzado nuestro objetivo.

La postura anti-realista que he expuesto en las últimas páginas permite defender la existencia de múltiples mundos reales que serían inconmensurables entre sí. Siendo así, podemos suponer que la solución al problema de la identidad personal en el NPN podría partir de la asunción de que los seres humanos, los Yoes y las personas son entidades que existen en mundos distintos, lo que explicaría que fuesen irreducibles una a la otra. Dado que cada mundo se basaría en la detección de una serie de patrones distintos, es no solo posible, sino también esperable que las condiciones de supervivencia de estas tres entidades sean independientes. Y como ninguno de estos mundos podría ser considerado como más «real» o «auténtico» que los otros, estaríamos plenamente justificados para identificarnos con las tres entidades y alegrarnos por su supervivencia en el mismo grado. De este modo, y según el tipo de patrones en los que nos estemos fijando para comprendernos a nosotros mismos, podríamos ser un ser humano físico, un Yo mental y una persona social al mismo tiempo en

distintos mundos. La confusión que dio origen al problema de la identidad personal procedería entonces del hecho de que en los tres tipos de mundo nos referimos a nosotros mismos con el pronombre personal «yo», lo que nos hace pensar que siempre nos referimos a una misma y única entidad que debe poseer todas las propiedades que tratemos de adscribirle. Sin embargo el hecho de que como vimos en el capítulo anterior las condiciones de supervivencia de estas tres entidades sean distintas demuestra que pese a las apariencias pertenecen a mundos diferentes, que todas ellas son descripciones válidas pero incompatibles de aquello que somos.

Para poder defender este punto de vista será necesario definir en qué consisten los tres mundos que estamos postulando, qué somos exactamente en cada uno de ellos y cómo es posible que vivamos al mismo tiempo en los tres. A ello dedicaré el próximo capítulo.

CAPÍTULO 5

EL NUEVO PLANTEAMIENTO NARRATIVO

En el segundo capítulo expuse las principales tesis del PNC, así como las críticas que el planteamiento ha recibido hasta la fecha. Tras ello apunté que la respuesta a dichas críticas debía partir de una diferenciación entre lo que somos como seres humanos, como Yoes y como personas, pues ésta parece ser la única forma de dar al mismo tiempo una explicación satisfactoria tanto de los aspectos fenomenológicos como sociales de nuestra existencia, tarea en la que, como hemos podido ver, tanto el planteamiento objetivo de la identidad personal como el narrativo clásico fracasaron. Sin embargo, antes de poder avanzar por dicho camino y explorar el NPN era necesario aclarar el marco ontológico en el que podría desarrollarse, pues de otro modo no podríamos defender que cuando nos preocupamos por nuestra supervivencia nos importan en igual medida las tres entidades que somos. Tras haber mostrado en los dos capítulos anteriores qué marco podría ofrecernos el sustento que necesitamos, ha llegado el momento de desarrollar el NPN, responder el resto de críticas que han recibido las tesis narrativas y elaborar mi postura con respecto al problema de la identidad personal.

Para ello, en primer lugar concretaré un poco más de qué modo el anti-realismo constituye una solución, mostrando cómo se crean los mundos en los que existimos como las tres entidades que nos ocupan y qué características poseemos en cada uno de ellos. Así veremos cómo los seres humanos existen en un plano de diseño en el que se realiza lo que generalmente consideramos propiedades físicas, los Yoes en uno intencional construido a partir de propiedades fenomenológicas y las

personas en uno colectivo formado por propiedades sociales. Asimismo abordaré las diferencias que existen en el NPN entre las propuestas de Dan Zahavi y Marya Schechtman con respecto a la diferenciación entre Yoes y personas.

En segundo lugar desarrollaré la idea de que dado el marco ontológico que hemos erigido no podemos establecer relaciones de dependencia entre estas tres entidades, exponiendo al mismo tiempo cómo el problema de la identidad personal surgió por la confusión producida al intentar adscribir las propiedades que tenemos en los distintos mundos a una única entidad. En este sentido, volveré sobre los experimentos mentales que motivaron el abandono del PO para mostrar cómo, desde la nueva versión de la identidad narrativa que propongo, éstos tienen una solución sencilla.

5.1. El mundo de los seres humanos

Si recordamos la analogía con el juego de la vida que expuse en el capítulo anterior, el mundo en el que existen los seres humanos se desarrollaría en el nivel de diseño, junto al resto de los objetos macroscópicos. Es un mundo formado por objetos físicos en el sentido popular del término, que no en el científico. Son objetos observables a simple vista y compuestos por distintos tipos de *materia*, elemento muy distinto de la materia/energía de la que habla la física cuántica. Dentro de este mundo aquello que son los seres humanos podría identificarse con la definición que la biología da del *homo sapiens* como especie animal. Aunque la biología actual podría sufrir una revolución en su seno que llevase a una nueva descripción de lo que es el ser humano, esto no sería un problema, pues si la descripción de lo que es un ser humano cambia, los seres humanos cambiaremos con ella. Como no nos hemos cansado de repetir a lo largo del capítulo anterior, ningún mundo subjetivo adquiere su existencia gracias a su correspondencia con el mundo REAL, sino por la detección y definición de patrones observables por nosotros en lo que quiera que la REALIDAD sea. De este modo el ser humano existe como aquello que podemos definir a partir de una serie de patrones físicos, visuales y táctiles principalmente, de comportamiento. Si la biología cambia su forma de entender e interpretar dichos patrones, los seres humanos cambiaremos también.

Un aspecto que merece una especial atención son los motivos que me llevan a inscribir al ser humano en el nivel de diseño en lugar de en el nivel micro-físico o subatómico, pese a la casi unánime creencia de que uno es reducible al otro. En primer lugar, aunque intentásemos hacer una

descripción de qué es lo que nos constituye en el nivel micro-físico en un determinado momento, dicha descripción sería poco útil, pues las partículas físicas que nos constituyen cambian constantemente a un ritmo frenético –de media un billón de átomos de nuestro cuerpo son reemplazados cada millonésima de segundo (Greene, 2004, p. 441)–, haciendo difícil concretar qué tipo de entidad cambiante somos si no pasamos a un nivel de diseño. Pero sobre todo el problema aparece cuando nos enfrentamos con una de las características más sorprendentes del universo: la existencia de la vida.

Pese a que desde nuestra perspectiva la diferencia entre un animal y una piedra es evidente, desde el nivel micro-físico la definición de qué es un ser vivo se torna extremadamente compleja, haciendo que las fronteras entre lo vivo y lo inerte se difuminen. Como ejemplo baste recordar la dificultad para determinar si entidades mucho más cercanas al nivel molecular que nosotros, como los virus, son un ser vivo o no. En la misma línea, parece imposible justificar por qué la vida ha de ser algo valioso cuando lo único que podemos describir en el mundo son diminutas cuerdas vibrantes de energía; qué hace importante que un grupo de cuerdas –por ejemplo las que forman los átomos de mi corazón– estén unidas entre sí o no. De este modo se hace evidente que si la vida de un ser humano es valiosa es porque se considera que nosotros, en tanto entidades existentes en un mundo del nivel de diseño, somos entidades con unas propiedades y una importancia que no se pueden justificar desde el nivel subatómico.¹

Otro aspecto sobre el que me gustaría detenerme brevemente para justificar mi definición de los seres humanos en un mundo del nivel de diseño, y que nos conecta inevitablemente con la definición de los Yoes en un mundo del nivel intencional, es el problema de la conciencia. Pues aunque los seres humanos no tienen conciencia en el sentido fenomenológico del término –esto es evidente dado que como mostré en el capítulo anterior las propiedades biológicas y las fenoménicas no pueden existir en un mismo mundo–, sí la tienen en el sentido psicológico. Siendo así, se hace necesario que el nivel en el que existen los seres humanos nos permita hablar de creencias, deseos, e incluso de sensaciones más primarias como las de dolor o placer, lo que podría inducirnos a pensar que los seres humanos pertenecen a un mundo del nivel intencional. En

1 A este respecto Fernando Broncano señala cómo la visión científica del mundo amenaza nuestra situación en él al no poder capturar la importancia de lo que nos preocupa, imponiendo una «Ley de la Realidad que vacía de absoluta significación cualquier acontecimiento, pues la vida o la muerte son palabras que para la realidad del universo físico no tienen sentido, puras descripciones de un continuo de causas-efectos» (Broncano, 2013, p. 99).

este sentido me gustaría señalar por un lado la dificultad de explicar no ya el sentido fenoménico, sino también este sentido psicológico de conciencia desde el nivel subatómico; y por otro la no necesidad de describir a los seres humanos desde el nivel intencional, pues la conciencia psicológica es explicable desde el nivel de diseño.

En primer lugar, resulta evidente y es ampliamente aceptada la dificultad que entraña explicar la existencia de propiedades fenoménicas hablando desde el nivel subatómico.² Sin embargo no sólo es difícil explicar la conciencia fenoménica en clave micro-física sino que también lo es explicar la conciencia psicológica. Pues del mismo modo que resulta difícil decir en qué consiste la vida cuando sólo se puede hablar de partículas subatómicas, la existencia de creencias o intenciones es difícil de comprender en clave de leptones o quarks. No resulta para nada intuitiva la idea de que billones de partículas dispuestas de tal y cual manera deban interpretarse como una creencia, cuando los movimientos de dichas partículas parecen ser perfectamente explicables en una clave mecanicista mucho más simple y efectiva sin necesidad de atribuirles una mente que por otro lado no queda claro a quién o a qué se le debería atribuir –¿A las partículas que forman los átomos de las neuronas implicadas? ¿Al conjunto de partículas que componen a un ser humano en un determinado momento? ¿Qué ocurre entonces con las partículas que abandonan nuestro cuerpo a cada segundo? ¿Siguen formando parte de la creencia aunque ya no formen parte del cuerpo?–.

En segundo lugar, como vimos en el capítulo anterior, resulta imposible compaginar en un mismo sistema las propiedades biológicas y las fenoménicas, lo que me llevó a defender que ambos tipos de propiedades pertenecen a mundos diferentes. Siendo así, al estar definiendo a los seres humanos a partir de sus propiedades biológicas resulta imposible tratar de adscribirles una fenomenología, lo que podría hacernos pensar que tampoco se les puede atribuir conciencia en el sentido psicológico. Sin embargo, de lo que expusimos en el capítulo tres se

2 Aunque no podemos dejar de señalar que no ha dejado de intentarse. Por un lado, como vimos en el capítulo tercero, los pansiquistas prometen una explicación de las propiedades mentales superiores a partir de la existencia de propiedades proto-mentales en las partículas físicas últimas de las que esté compuesta la REALIDAD, aunque para ello deben primero solventar el problema de la combinación, lo cual parece lejos de poder alcanzarse. Por otro, algunos autores como Roger Penrose han tratado de explicar la conciencia como un producto de efectos cuánticos en el nivel neuronal, aunque se trata de una posibilidad improbable que no ha tenido una gran acogida por parte de la comunidad científica o filosófica (*vid.* Penrose, 1989, 1994).

recordará que existe una postura que trata de explicar las propiedades mentales de los seres humanos en el plano psicológico negando que además haya una conciencia fenoménica: el materialismo eliminativo. Los eliminativistas no niegan que los seres humanos tengan dolores, creencias o deseos. Lo que niegan es que exista una dimensión fenomenológica, no-física, de los mismos. De este modo, consideran que las propiedades mentales existen, pero son reducibles a propiedades neuronales, esto es, a propiedades del nivel de diseño, y son por tanto comprensibles dentro del mismo esquema causal físico que nos permite comprender el resto de objetos macroscópicos. De este modo dentro del mundo en el que existen los seres humanos podemos atribuirles una mente y una conciencia, aunque éstas deben ser entendidas en el sentido en que las entienden los eliminativistas.³

Resumiendo lo anterior podemos decir que los seres humanos son entidades existentes en un mundo del nivel de diseño que poseen propiedades no explicables desde el nivel físico subatómico, que están compuestos totalmente por materia en un sentido más newtoniano que cuántico y cuyo comportamiento se explica por la posesión de propiedades mentales, aunque éstas deben entenderse como absolutamente reducibles al funcionamiento neuronal. En cuanto a su identidad, se trata de una relación indeterminada, pues como vimos en el primer capítulo con respecto a los animalistas, nuestro cuerpo se puede ver sometido a situaciones en las que sea imposible decidir si ha sobrevivido o no. En cualquier caso volveré brevemente sobre esto.

5.2. El mundo de los Yoes

Las características del mundo en el que existen los Yoes fueron también presentadas en el capítulo anterior, pues se trata de un mundo creado por la actitud intencional definida por Daniel Dennett y cuyas guías maestras son los rasgos de la psicología del sentido común. Sin embargo, como se recordará, en el tercer capítulo señalé que la diferenciación entre Yoes y personas tenía un cariz distinto en las posturas de Zahavi y

3 A este respecto es interesante la posición de José Luis Bermúdez, quien intenta hacer una reducción del Yo a un marco fisicalista. Aunque consigue mostrar en qué sentido sería posible explicar la dimensión psicológica de los pensamientos en primera persona en términos físicos, la dimensión fenomenológica de los mismos se encuentra totalmente fuera de su radar. Sencillamente es un aspecto que ni siquiera puede plantearse dentro de un marco reductivo (*vid.* Bermúdez, 1997).

Schechtman, por lo que era necesario hacer alguna aclaración al respecto. Aunque ya anteriormente señalé mi inclinación hacia la postura de Schechtman, ha llegado el momento de justificar mi posición para poder distinguir adecuadamente entre ambos conceptos y definirlos con una precisión mayor.

5.2.1. Fenomenología y narratividad

Como se recordará, según Zahavi un enfermo de Alzheimer avanzado incapaz de recordar ni siquiera lo que le acaba de suceder puede experimentar como propio un dolor sin necesidad de recurrir a ningún tipo de narración. Sin embargo para comprender nuestra vida extendida en el tiempo hemos de hacerlo en forma narrativa, por lo que se hace necesario distinguir entre un concepto fenomenológico, mínimo, de Yo, y otro concepto más extendido, narrativo, al que para diferenciarlo del Yo podemos denominar «persona». De este modo según Zahavi el Yo estaría relacionado con un nivel de conciencia fenomenológica bruta, existiría de forma previa a toda narración, y sólo en los casos en los que alguien construye un Yo extendido o persona las narraciones tendrían algún papel, aunque éste sería de desarrollo, no constitutivo (Zahavi, 2007, pp. 200-201). Sin embargo para Schechtman el Yo, aunque también se vincula al modo en que subjetivamente experimentamos nuestra vida, está constituido por narraciones. Aunque éstas no tienen por qué tener una duración temporal extensa, si alguien no tiene una comprensión narrativa de su vida en tanto Yo entonces no es en absoluto un Yo, pues para constituirse como tal es necesario tener una narración que unifique todos los episodios que le pertenecen (Schechtman, 2007, p. 171). Por su parte las personas estarían también constituidas narrativamente, aunque su estructura narrativa estaría más desarrollada. En cualquier caso me ocuparé de las personas en el siguiente apartado. De este modo, y centrándonos en el problema de los Yoes, a pesar de que tanto Zahavi como Schechtman aceptan la distinción entre ambos tipos de entidades y por tanto el NPN, siguen discrepando acerca del sentido de esta diferenciación, pues lo que Schechtman llama «Yo» se parece a lo que Zahavi llama «persona», de manera que da la impresión de que cuando Zahavi habla de «Yo» se refiere a una entidad todavía más primitiva que el Yo de Schechtman.

Schechtman sugirió dos posibles líneas de respuesta –que en el fondo están interrelacionadas– que harían compatibles la postura de Zahavi con la suya, aunque en mi opinión no acierta a desarrollarlas correctamente

(Schechtman, 2011, pp. 410-411). En su reporte de la postura de Zahavi, Schechtman sugiere que podría haber varios significados del término «Yo», y que Zahavi y ella simplemente podrían estar refiriéndose a significados distintos, lo que eliminaría cualquier posible discusión. Por otro lado también sugiere que el tipo de conciencia fenomenológica bruta que compartimos con los animales de la que habla Zahavi podría no requerir de un carácter narrativo, mientras que el tipo de auto-conciencia que poseemos nosotros y a la que ella se refiere sí la requeriría. De este modo parece indicar que la conciencia fenomenológica bruta da lugar a Yoes no-narrativos y la auto-conciencia típica de los humanos forma Yoes narrativos. Sin embargo, aunque en líneas generales creo que su respuesta es adecuada, como ella misma reconoce este movimiento hace que tenga problemas para dar un lugar apropiado a la dimensión fenomenológica de nuestra existencia en su teoría, lo cual podría volver a llevarla al mismo punto que llevó al colapso tanto al PO como al PNC (Schechtman, 2011, pp. 411, 415). Por otro lado, considero que la diferenciación entre estos dos tipos de Yoes para dar cabida a la dimensión fenomenológica de nuestra existencia supone una multiplicación innecesaria de entidades, pues como vimos en el segundo capítulo distinguir entre Yoes y personas ya tenía como uno de sus objetivos precisamente esta cuestión. Para solucionar este problema debemos por tanto elaborar un concepto de «Yo» narrativo que tenga en cuenta su carácter fenoménico, de modo que si Zahavi se quiere referir a un significado distinto de «Yo», la diferencia tenga que estar en algún otro sitio.

El problema desde mi punto de vista reside en compaginar dos aspectos del Yo en los que Zahavi y Schechtman hacen distinto énfasis: su existencia como sujeto de experiencias y como agente auto-consciente (*vid.* Korsgaard, 1989). Al hacer hincapié en el Yo como sujeto de experiencias, Zahavi (igual que Strawson) parece reducir su existencia a la de simple entidad sintiente. Por su parte Schechtman, más preocupada por nuestra vida activa, resalta la dimensión del Yo como agente auto-consciente. De este modo mientras que Zahavi, preocupado por la vertiente más pasiva de los Yoes, puede prescindir de la noción de narratividad, para explicar su condición de agentes Schechtman necesita recurrir a ella. Pues un agente para ser auto-consciente debe comprender sus acciones de un modo adecuado, y para comprenderlas debe hacerlo de un modo narrativo. El problema previo que deberíamos resolver aquí, pero al que todavía no me quiero enfrentar, es el de qué estamos entendiendo por «narración».⁴ En cualquier caso trataré de dar una respuesta provisional que será compatible con lo que discutiré más adelante.

4 Le dedicaré mi atención a este problema en el sexto capítulo.

Schechtman recurre a un concepto de narración bastante formalista, lo cual le impide dar un lugar adecuado a la conciencia fenoménica. Para Schechtman los Yoes están constituidos de forma narrativa en el sentido de ser capaces de dar cuenta explícita de forma verbal de sus acciones conectándolas entre sí. Es decir, se constituyen como Yoes en la medida en que son capaces de producir un discurso sobre su vida con los rasgos formales de una narración (Schechtman, 2007, pp. 159-160). De este modo el poseer la capacidad de hablar un lenguaje se convierte para Schechtman en un requisito indispensable para ser un Yo, mientras que Zahavi señalaría acertadamente que no saber expresar lingüísticamente cómo es sentir una sensación de dolor no implica que no exista un Yo que experimente conscientemente dicha sensación como propia. Sin embargo, aunque coincido con Zahavi en que no es necesario el lenguaje para experimentar algo de forma consciente, sí creo que es necesaria una cierta comprensión de nuestra vida mental para ser un Yo. Al menos un Yo normal de tipo humano. En este sentido, aunque no es necesario tener unas categorías desarrolladas como las que posibilita la posesión de un lenguaje articulado, creo que es necesario para ser un Yo tener la capacidad de representarse y comprender la propia vida mental. Es decir, es necesario ser autoconsciente no en el sentido lingüístico de Schechtman, sino en el de tener la capacidad de transformar la interacción con el mundo en *experiencia*, donde experiencia debe entenderse en la forma en que Broncano desarrolla el concepto:

Denominamos «experiencia» a una interacción con algo que se impone, que en un sentido nos sobrepasa, pero que al tiempo es asumido como cambio que transforma al sujeto. Puede ser una experiencia observacional o perceptiva, tal vez propioceptiva. Pueden ser experiencias en un sentido lato que implican procesos amplios, experiencias que abarcan periodos de la vida. En cualquier caso, por experiencias nos referimos a *unidades de sentido*, trozos de vida que son comprensibles o que producen comprensión. El concepto de experiencia abarca un dominio más amplio que el que suele considerar la filosofía analítica, donde la experiencia queda reducida y tratada bajo el epítome de *percepción*, de modo que todo su complejo de connotaciones se ocluye bajo su dimensión puramente epistémica de indicador fiable a través de un espacio de *sensibilia* del que apenas importa otra cosa que su carácter cualitativo y fenoménico. (Broncano, 2013, p. 50)

La experiencia así entendida incluye la dimensión fenoménica subjetiva pero no se limita a ella, pues presta atención así mismo a otros dos aspectos que van más allá de la mera «percepción analítica». En primer lugar, tiene en cuenta que todas las experiencias –incluso las

propioceptivas– implican una cierta interacción con el mundo exterior, pues los Yoes no son simples sujetos pasivos, sino que son agentes que reaccionan públicamente ante las experiencias y adoptan una actitud proactiva hacia ellas. En segundo lugar, toma en consideración que las experiencias influyen en los Yoes de un modo duradero, pues al dotarlas de sentido e interiorizarlas se produce un aprendizaje y en consecuencia una evolución del propio Yo (Broncano, 2013, pp. 51-52).

Ser capaz de tener experiencias en este sentido creo que es independiente de la capacidad para tener un lenguaje articulado y por tanto de la de producir narraciones en el sentido formal del término utilizado por Schechtman. O en todo caso creo que la primera podría ser una condición necesaria pero no suficiente para la segunda. Pues por ejemplo, a pesar de que entre las especies que habitan la Tierra sólo los seres humanos parecen capaces de tener un lenguaje, el comportamiento de algunos animales parece sugerir que son auto-conscientes en el sentido que acabo de señalar. Así interpreto que ciertas especies de monos sean capaces de diferenciar entre aquello que está presente en su mente y lo que está presente en la mente de los otros y de utilizar este conocimiento para el engaño. Esto es, creo que para que este comportamiento sea posible es necesario que los monos comprendan sus propias creencias, deseos e intenciones, así como los de sus rivales, provoquen situaciones de forma activa que lleven a estos últimos a comportarse del modo que desean y aprendan de sus errores pasados para perfeccionar sus técnicas de manipulación.⁵

5 Es importante señalar que Broncano rechaza que los animales puedan tener experiencias, pues pone una carga mucho mayor en lo que la involucración con el mundo implica. Según Broncano para que algo sea una experiencia no basta con que tenga una expresión pública, sino que tiene que ser reconocido socialmente como un cambio apreciable en la trayectoria vital de un miembro del grupo, lo cual sólo sucede en el caso de los seres humanos (Broncano, 2013, p. 52). Por otro lado, que los monos sean realmente auto-conscientes no es algo ni mucho menos aceptado como evidente por la comunidad científica, sino que es algo abierto a discusión. En cualquier caso este aspecto no afecta realmente a mi exposición, pues mi intención no es defender que los monos puedan tener Yoes humanos, sino que la capacidad de tener experiencias puede darse en ausencia de lenguaje. Según mi postura los Yoes humanos tienen una capacidad de transformar su interacción con el mundo en experiencia mucho más desarrollada que la de los Yoes de los monos, por lo que existe una diferencia clara entre ambos. Y aunque los monos no fuesen auto-conscientes, esto no implicaría que el lenguaje fuese necesario para tener experiencias, sino únicamente (quizá) que la diferencia entre nuestros Yoes y los de otros animales es una cuestión de cualidad y no de grado, como aquí estoy presuponiendo.

El concepto de experiencia que estoy empleando nos permite dar cabida a nuestra dimensión de sujetos –lo cual constituye un avance con respecto a la postura de Schechtman–, al mismo tiempo que atiende a nuestra faceta de agentes. De este modo, al unificar ambas dimensiones bajo un único concepto de Yo, mi postura permite enfrentar el caso de los enfermos graves de Alzheimer señalando que los Yoes humanos pueden deteriorarse hasta el punto de no tener las funciones básicas para ser considerado un Yo normal, sin que ello implique la necesidad de postular un sentido distinto de Yo para atender a la dimensión fenomenológica (aunque si Zahavi insistiese en ello podríamos otorgarles un reconocimiento específico a este tipo de Yoes deteriorados). Siendo así, presuponiendo que la capacidad de tener experiencias en el sentido indicado está al alcance de todos los Yoes humanos que funcionen con normalidad, y que por tanto éstos además de tener conciencia fenoménica poseen unas ciertas categorías mentales pre-lingüísticas para poder comprenderse a sí mismos,⁶ podemos recurrir a la definición de narración de Bruner, quien como ya hemos dicho considera que nuestro pensamiento es narrativo cuando se realiza en términos intencionales, para asegurar que el pensamiento de los Yoes sobre sí mismos es necesariamente narrativo. Siendo así considero que los Yoes están constituidos narrativamente, pues para ser un Yo es necesario poder pensar sobre uno mismo en términos intencionales. O lo que es igual, un Yo sólo es un Yo si piensa narrativamente sobre sí mismo. Evidentemente Strawson consideraría trivial esta afirmación, pero me ocuparé de esta acusación más adelante.

Teniendo en mente lo anterior, podemos completar la definición de «Yo» que dimos en el tercer capítulo afirmando que un Yo es una entidad mental caracterizada por ser el sujeto de las experiencias conscientes, ser un agente racional, tener autoconciencia y tener cierto tipo de personalidad que le hace tomar decisiones basadas en las sensaciones que el contacto con el mundo exterior le provoca, en las creencias que posee y en los deseos que se producen en la interacción entre sensaciones y creencias. Esta adscripción de «personalidad» de nuevo querría ser denegada por Zahavi, quien asegura que los Yoes no poseen características que los individualicen de un modo significativo (Zahavi, 2007, p. 193), sin embargo creo que el concepto de «Yo» que estamos desarrollando implica una cierta diferenciación, pues cada Yo entenderá de un modo

6 Esta idea parece sugerir que la *folk psychology* podría tener una base biológica que por tanto sería universal, lo cual es una tesis discutible que merecería mayor atención (*cfr.* Hutto, 2007). En cualquier caso en el sexto capítulo afinaré la definición de qué significa entenderse a uno mismo en forma narrativa.

característico su propia vida mental. Por otro lado, y como ya dijimos capítulos atrás, debemos considerar que un Yo es una entidad distinta del cuerpo que habita, aunque no posee independencia con respecto al mismo, pues debe estar encarnado en el cuerpo concreto que lo genera y sustenta. Sin embargo es importante notar que un Yo no es un cuerpo, sino que posee un cuerpo del que no se puede separar.⁷

En este momento alguien podría caer en el error de confundir el cuerpo en el que habitan los Yoes con el ser humano que definimos en el apartado anterior, pero se trata de dos entidades distintas que existen en dos mundos distintos. Los cuerpos no necesitan estar vivos para seguir existiendo –un cuerpo no deja de ser un cuerpo por ser cadáver, si acaso se hace más evidente su condición–, mientras que los seres humanos dejan de existir al fallecer. Por otro lado los seres humanos, como acabamos de ver, poseen cualidades mentales, mientras que los cuerpos carecen de ellas. O en todo caso sólo las poseen de forma vicaria por su relación con el Yo que encarnan (Parfit, 2012, pp. 23-24). Esto es así porque los Yoes por definición poseen las cualidades mentales tanto en su dimensión fenomenológica como en la psicológica, por lo que no es necesario tratar de adscribirlas a sus cuerpos.⁸

7 Me sirvo en gran medida del análisis que Strawson hace de la naturaleza de la experiencia del Yo, aunque discrepo de sus conclusiones. Según Strawson, en la experiencia corriente del Yo se considera a éste como (1) un sujeto de experiencias que es (2) una cosa (3) simple, (4) persistente, (5) mental que es (6) un agente que tiene (7) una cierta personalidad y que (8) no es idéntico al ser humano tomado en su totalidad. De estas ocho características, Strawson defiende que en realidad sólo son necesarias (1), (2), (3) y (5): ser un sujeto de experiencias que es una cosa simple de naturaleza mental (SESMET, por sus siglas en inglés). Por mi parte considero que también son ciertas (6), (7) y (8), y que además es importante el requisito de que tengan conciencia de sí mismos. El problema de (4), la persistencia a lo largo del tiempo, es que es variable según los casos, como explicaré un poco más adelante (*cf.* Strawson, 2009, pp. 161-216).

8 Alguien podría notar en este momento la dificultad de explicar entonces el aspecto físico de la existencia de los Yoes, su relación con el cerebro o el cuerpo que habitan, desde este mundo intencional. Se trata de una dificultad –o imposibilidad– análoga a la que nos encontramos en el mundo de los seres humanos para explicar la existencia de la conciencia fenomenológica. Como no nos hemos cansado de repetir, las propiedades biológicas y las fenomenológicas no pueden existir en un mismo mundo, motivo por el cual al centrarnos ahora en la existencia de los Yoes nos encontramos con problemas a la hora de explicar el mundo físico. Sencillamente no es posible ofrecer una explicación satisfactoria de todos los aspectos al mismo tiempo.

5.2.2. *Supervivencia e identidad*

Otro aspecto que merece la pena discutir es cuáles son las condiciones de supervivencia de los Yoes y cuál es su duración típica, pues como vimos en el segundo capítulo una de las principales reivindicaciones del PNC era que un Yo normalmente tiene una duración temporal idéntica a la de la vida del cuerpo humano que le encarna, y que para adquirir esa duración tiene que tener una comprensión narrativa de si mismo. Sin embargo esta premisa fue severamente criticada por Strawson basándose en la defensa de que existen personas Episódicas para las que dicha afirmación es completamente falsa. Siendo así se hace necesario aclarar este aspecto para poder continuar desarrollando el NPN.

En primer lugar podemos ver que evidentemente, una vez que hemos diferenciado entre Yoes y personas, la necesidad de postular un Yo con una duración tan extendida ya no es tan apremiante, pues como vimos al analizar el PNC esta necesidad de presentar toda nuestra vida unificada en un único relato procede de la imputación de identidad estricta que demanda nuestra existencia social. Dado que dichas demandas sociales se deben dirigir ahora hacia las personas, no parece existir ninguna presión en este sentido para que otorguemos a los Yoes una duración mayor. En segundo lugar podemos darnos cuenta de que tal y como lo estamos definiendo existe un acceso privilegiado al Yo desde una perspectiva de primera persona, y por tanto cada Yo tiene la última palabra sobre su propia existencia. Si alguien dice que ya no es el mismo Yo que habitaba su cuerpo el día anterior, nadie puede presentar un argumento en contra que pueda rivalizar con la percepción subjetiva de ese mismo Yo.⁹ Siendo

9 El caso contrario, el de alguien que asegura ser el mismo Yo que era el día anterior, presenta complicaciones adicionales, poniendo de evidencia que a pesar de que el Yo tenga un acceso privilegiado desde la primera persona, existen interpretaciones válidas e inválidas también en los mundos del nivel intencional. Si recordamos los casos de duplicación que presenté en el primer capítulo, seguramente alguien que poseyese una copia idéntica de la mente de Napoleón aseguraría ser el mismo Yo que habitaba el cuerpo de Napoleón antes de morir. Sin embargo como vimos en aquel momento esto no es posible. Incluso aunque no fuese un simple demente y su mente fuese efectivamente una copia exacta de la de Napoleón, no sería el mismo Yo, pues no podría poseer las mismas experiencias conscientes si éstas son debidas a una base física distinta –recordemos que los Yoes no pueden separarse del cuerpo que los encarna–. El «Napoleón» de nuestro experimento debería aceptar esto para poder construir un mundo válido. En cualquier caso, el experimento pone de manifiesto otro problema al que no prestaré atención aunque es importante señalar, y es el de hasta qué punto alguien puede estar confundido sobre su propio Yo.

así, las reivindicaciones de Strawson sobre la existencia de personas Episódicas, poniéndose a sí mismo como ejemplo, no pueden ser puestas en duda.

Estas consideraciones creo que serán compartidas por todos los participantes en el debate y pueden ser consideradas parte integral del NPN. Sin embargo no por ello todos aceptan que debemos resignarnos a que cada Yo dure lo que tenga que durar, pues como dije en el segundo capítulo Schechtman señala que deberíamos adoptar como un deber moral el compromiso de intentar que la duración de nuestro Yo sea tan extensa como la duración de la persona y del ser humano que también somos (Schechtman, 2007, pp. 175-178). Para ello alega dos motivos. El primero es que si no lo hacemos podrían producirse problemas afectivos y éticos, pues considera que tener un Yo cuya duración no coincida con la duración de la persona que también somos es una receta perfecta para la alienación. El segundo es que los Yoes de los que estamos hablando desde una perspectiva más o menos fenomenológica nunca están tan claramente definidos como Strawson parece sugerir al hablar desde una perspectiva más metafísica. Personalmente no siento ninguna necesidad de establecer relaciones entre los tres tipos de entidades que somos, ni siquiera en el sentido sugerido por Schechtman de que deberíamos procurar que sus duraciones fueran coextensivas. Sin embargo, para aclarar qué es lo que considero que somos como Yoes puede ser de utilidad tratar de responder a sus afirmaciones.

En primer lugar, Schechtman asegura que si la duración del Yo no es tan amplia como la duración de la persona inevitablemente deberemos enfrentarnos a situaciones en las que nuestra existencia social nos demande un comportamiento contrario a nuestras emociones subjetivas. En un ejemplo que ella misma reconoce poco elaborado asegura que aceptar esto sería como reconocer que debemos cuidar de nuestros hijos porque la sociedad nos lo exige pero subjetivamente no sentir ningún afecto hacia ellos ni tener ninguna preocupación por su bienestar, pues en cierto sentido podríamos pensar que en realidad no son nuestros hijos, sino los hijos de un Yo que murió tiempo atrás. Este tipo de situaciones, dice Schechtman, inevitablemente deberían hacernos sentir alienados, por lo que para evitarlas es preferible hacer todo lo posible para que nuestro Yo extienda su duración a lo largo de toda la vida del ser humano que también somos.

El problema de esta crítica es que presupone que sólo podemos identificarnos en un alto grado con una de las entidades que somos, el Yo. De este modo aunque en cierto sentido podamos reconocer como «nuestros» episodios de nuestra vida (pasados o futuros) que nos

pertenecen o pertenecerán en tanto personas o en tanto seres humanos, no serán real o completamente *nuestros* a no ser que también nos pertenezcan en tanto Yoes. Esto es así porque el Yo queda definido para Schechtman de tal modo que cualquier episodio con el que tengamos una vinculación emocional forma parte de nuestro Yo, monopolizando esta entidad toda nuestra vida afectiva. Así, frente a los episodios con los que sentimos alguna vinculación, todos los demás aspectos del mundo –incluidos los que nos involucran a nosotros en tanto personas o en tanto seres humanos– parece que deben ser sentidos según la teoría de Schechtman como radicalmente ajenos. Sin embargo si esto fuese cierto la tasa de suicidios entre los Episódicos debería ser del 100%, pues partiendo del supuesto de que todo el mundo se siente triste en algún momento, si un Episódico no sintiese ninguna vinculación emocional hacia su futuro en tanto ser humano debería sentirse movido a poner fin a su sufrimiento (y a su existencia) de forma inmediata, pues al fin y al cabo su vida en tanto Yo (que sería la única que le importaría) apenas iría a durar unos segundos más. Es decir, si sólouviésemos una vinculación emocional con lo que somos como Yoes, deberíamos actuar en consecuencia y ser mucho más egoístas en lo que se refiere a nuestra relación con lo que somos como personas y seres humanos. Si esto no sucede no es porque los Episódicos en realidad no sean tan Episódicos como dicen, sino porque el Yo no monopoliza nuestra vida afectiva, como parecen sugerir Schechtman y Strawson.¹⁰ Aunque haya sujetos que puedan identificarse en un grado mucho más alto con lo que son como Yo que con lo que son como persona o como ser humano, la mayoría de nosotros se identifica en distintos grados al mismo tiempo con las tres entidades. Y poniéndome a mí mismo como ejemplo puedo asegurar que es posible identificarse con las tres en la misma medida. De este modo, aunque no me identifico con el Yo que era cuando empecé a trabajar en esta tesis ni con el que disfrutará del grado de doctor en el futuro, sí me identifico con la persona y con el ser humano que era y que seguiré siendo, de modo que no siento ningún tipo de alienación por tener que continuar mi propio proyecto, aunque mi Yo actual hubiese preferido hacer otra cosa. Es decir, al identificarme con las

10 En realidad el problema de Strawson es que se centra demasiado en nuestra existencia como sujetos de experiencias olvidando nuestra dimensión de agentes, pues como Christine Korsgaard mostró la idea de agencia requiere que el Yo tenga una existencia temporal más amplia y explica por qué nos preocupamos por nuestro futuro y no únicamente por nuestro estado presente (*cf.*: Korsgaard, 1989, pp. 126-127; Strawson, 2007). Por su parte Schechtman, como ya hemos visto, sí tiene en cuenta que somos agentes, pero le da demasiada importancia a la vinculación emocional.

tres entidades a la vez puedo manejar las distintas metas y deseos que tengo, frustrando algunas en beneficio de otras sin necesidad de dar preeminencia en todo momento a los deseos que beneficien a una de las tres entidades.¹¹

El segundo motivo que lleva a Schechtman a defender la necesidad de prolongar lo más posible la existencia de nuestro Yo es que los Yoes fenomenológicos no están tan claramente divididos como los Yoes metafísicos de los que habla Strawson. Para Strawson los Yoes se suceden unos a otros en nuestro cuerpo como las perlas en un collar (*vid.* Strawson, 1997, 1999), de modo que da a entender que un acontecimiento del pasado no puede ser sentido como propio a no ser que también sintamos como propio todo el periodo que media entre aquel momento y la actualidad. En consecuencia una vez que ya no sentimos como propio un periodo de nuestro pasado, todo lo que sucediese durante o antes de ese periodo debe ser sentido como perteneciente a otro Yo. Sin embargo señala Schechtman que nuestra vida no funciona así. Nuestros Yoes son más «fluidos» y «amorfos» (Schechtman, 2007, p. 177) y podemos sentir una vinculación emocional fuerte con un suceso pasado de nuestra vida a pesar de no sentirla hacia sucesos más recientes. Al mismo tiempo siempre es posible que un acontecimiento pasado hacia el que en la actualidad no sentimos ningún afecto recobre importancia para nosotros en el futuro. En este sentido Schechtman señala que deberíamos dar la oportunidad a nuestras características pasadas de expresarse en el presente mediante la extensión de nuestros Yoes, pues puede que pese a nuestro actual desinterés en realidad sean aspectos de nuestra vida que todavía nos pertenecen en tanto

11 A este respecto puede confrontarse mi postura con la que sostiene Harry Frankfurt acerca de los contenidos mentales con los que no nos identificamos de forma fuerte. Señala Frankfurt que en nuestra mente pueden existir pensamientos que preferiríamos no tener, que nos «asaltan» sin que los evoquemos (no implican una actitud activa por nuestra parte) y que consideramos ajenos a nuestro verdadero ser. Según su teoría es preferible considerar que estos pensamientos no nos pertenecen realmente. Del mismo modo que si una causa externa mueve nuestro cuerpo contra nuestra voluntad no podemos decir que dicho movimiento sea nuestro aunque sea un movimiento de nuestro cuerpo, los pensamientos que se sienten como ajenos deberían considerarse como externos a nosotros, aunque tengan lugar en nuestro cuerpo (*cf.* Frankfurt, 1976). Sin embargo creo que este tipo de discurso, en parte coincidente con el de Schechtman y Strawson, puede provocar una cierta enajenación, por lo que es preferible considerar que todos los pensamientos que se encuentran en nuestra mente nos pertenecen, pues aunque no nos identifiquemos con ellos en tanto Yoes, podemos hacerlo como personas o como seres humanos.

Yoes pero que se encuentran en estado latente, a la espera de volver a atraer nuestra atención.

Otra vez el problema de la postura de Schechtman es que define los Yoes de tal forma que monopolizan nuestra vida afectiva, de modo que aquello por lo que sentimos o podemos sentir algún afecto es parte de nuestro Yo y todo lo demás es algo ajeno. Sin embargo, aún reconociendo que los Yoes fenomenológicos son mucho más fluidos y amorfos que los Yoes metafísicos de Strawson, esto no implica que nuestro Yo deba extenderse en el tiempo para abarcar todos los episodios con los que sentimos una vinculación emocional. Por ponerme de nuevo a mi mismo como ejemplo, aunque siento una vinculación afectiva hacia mi vida en el periodo en que empecé el doctorado, recuerdo con cariño a los amigos que tenía entonces –muchos de los cuales todavía conservo– y reconozco su importancia en la formación de mi carácter y en mi situación actual, no por ello me identifico en tanto Yo con el Yo que era entonces aunque sí me identifique con el ser humano y con la persona que era.

En este momento, y antes de seguir adelante, se hace necesario abordar el principal problema de mi caracterización del Yo, que lleva exigiendo una respuesta desde que aseguré que éste no monopoliza nuestra vida afectiva y que podemos (o no) identificarnos por tanto como cualquiera de las tres entidades que somos con etapas a las que estamos vinculados emocionalmente. Por plantearlo de forma explícita: si tengo una cierta vinculación emocional tanto con mi vida pre-doctoral como también con lo que me sucedió ayer, y me identifico en ambos casos con alguna de las entidades que era entonces, ¿cuál es la diferencia entre identificarse en tanto Yo e identificarse en tanto persona o en tanto ser humano? ¿Qué es lo que me mueve a decir que aunque tengo una vinculación emocional con mi vida pre-doctoral y me identifico con el ser humano y la persona que era hace cuatro años, no me identifico con el Yo que era entonces? ¿Qué falta en la relación que hay entre mi Yo actual y mi Yo pre-doctoral que sin embargo si está presente en la relación entre mi Yo actual y mi Yo de ayer si en ambos casos hay un vínculo afectivo y de identificación?

De mi postura se extrae como evidente que dado que todos somos al mismo tiempo tres entidades distintas, existen otras tantas formas en las que podemos identificarnos con quienes éramos en una etapa pasada. De este modo, dadas unas coordenadas temporales pasadas en las que nosotros existíamos, puede suceder que nos identifiquemos de forma independiente (1) en tanto seres humanos con el ser humano que éramos, (2) en tanto Yoes con el Yo que éramos y (3) en tanto personas con la persona que éramos. Dado que las condiciones de supervivencia de estos tres tipos de entidades son distintas, entre la actualidad y cualquier

momento del pasado pueden darse sólo un tipo de identificación, dos cualesquiera de ellos, o los tres al mismo tiempo. En el ejemplo que estoy utilizando, entre mi vida actual y mi vida pre-doctoral sostengo que siento identificaciones de los tipos 1 y 3, pero no del tipo 2; por su parte entre mi vida ayer y mi vida hoy siento identificaciones de los tres tipos. El motivo que me lleva a considerar que no soy el mismo Yo que era cuando empecé el doctorado es que, aunque siento una vinculación afectiva con aquel momento de mi vida, las ideas, proyectos y creencias que tenía entonces son totalmente distintos de los actuales. No en vano mi proyecto inicial de tesis tenía que ver con la representación de la memoria en el cine posmoderno y no con el problema de la identidad narrativa en la filosofía analítica contemporánea. Es más, en aquel momento habría despreciado un tema así por considerarlo demasiado concreto –en cierta forma me parecía a ese estudiante de doctorado del que habla Umberto Eco que quería hacer una tesis titulada *La literatura hoy* (Eco, 1977, p. 27)–. En este sentido, a pesar de que aprecio realmente aquella etapa de mi vida, en líneas generales su recuerdo no afecta a mi toma de decisiones actual, pues mis intereses y creencias han cambiado tanto que no puedo conectarlas entre sí estableciendo una continuidad (*vid.* Schechtman, 2001). Desde mi perspectiva, sólo son parte de nuestro Yo aquellas etapas que son sentidas como propias de una forma más directa, con las que se puede establecer una continuidad respecto a nuestros deseos, creencias e intereses actuales y que por tanto afectan a nuestra toma de decisiones actual. Siendo así, considero que pueden existir vínculos emocionales y de identificación hacia otras etapas de nuestra vida sin que eso automáticamente signifique que esa identificación tenga que ser del tipo 2.¹²

El problema de las identificaciones de los tipos 1 y 3 en realidad tiene menos importancia que el de tipo 2, pues a diferencia de los Yoes, los seres humanos y las personas no se definen (o no exclusivamente) desde la primera persona. En este sentido alguien es o no es el mismo ser humano si los patrones físicos que podemos detectar en el mundo así lo indican,

12 Esta postura en realidad no está tan alejada de la Schechtman, pues para definir el Yo estoy tomando prestado su concepto de «acceso empático». El acceso empático es el tipo de acceso que uno tiene a los recuerdos de sus propios estados mentales pasados de forma que, pese a no ocurrir en el presente de la experiencia, se recuerdan como si le hubiesen ocurrido al mismo Yo que existe en el presente, afectando por tanto a su toma de decisiones (Schechtman, 2001, pp. 101-102). En lo que discrepo con Schechtman es con sus afirmaciones posteriores posicionándose a favor de que debemos extender nuestro Yo en las que además parece estar implícita la idea de que el Yo monopoliza nuestra vida emocional, como ya he señalado.

independientemente de las creencias al respecto del ser humano implicado. Del mismo modo, alguien es la misma persona si los patrones sociales que la definen lo determinan de ese modo, no importando (o importando en menor medida) lo que la persona pueda pensar. A este respecto puede pensarse en el caso del protagonista de *No soy Stiller* al que me referí en el capítulo dos, a quien el estado de Suiza lleva a juicio para obligarle a aceptar la «personeidad» de Stiller. En cualquier caso volveré más adelante sobre el papel que tienen las creencias de las personas en la constitución de su propia identidad. En este momento lo único que quiero señalar es que nos podemos identificar en tanto personas y en tanto seres humanos con nuestra existencia pasada mediante el reconocimiento de los patrones que nos definen como cada una de estas dos entidades y que dicha identificación puede ir acompañada de fuertes vínculos emocionales sin que eso cause ningún problema a nuestra consideración de lo que somos en tanto Yoes.

Volviendo sobre el tema de cuáles son las condiciones de supervivencia de los Yoes y cuál es su duración típica, ya hemos aclarado que un Yo sobrevive mientras alguien siga afirmando de una forma válida ser el mismo Yo.¹³ Al mismo tiempo hemos mostrado que en consecuencia su duración es flexible, ya que a pesar de que en tanto sujetos de las experiencias conscientes no necesitan tener una duración temporal mayor que la de una experiencia aislada o lo que Strawson denomina el «presente de la experiencia», cuya duración cifra en apenas unos milisegundos (Strawson, 2009, pp. 261-264), gracias al acceso empático que todos sentimos en forma diferente hacia nuestros estados mentales pasados los distintos Yoes pueden extenderse en el tiempo de modo variable, incluso pudiendo sobrevivir durante cientos de años como cerebros dentro de una cubeta (*vid.* Schechtman, 2001, pp. 101-102). De este modo podemos ver que en un mismo cuerpo pueden habitar distintos Yoes, ya sea de manera sincrónica –en los casos de personalidad múltiple– o diacrónica. El único requisito para que esto sea así es que subjetivamente alguien experimente su vida mental de este modo, comportándose en consecuencia.

En cualquier caso un tema importante que todavía resta por aclarar a este respecto está relacionado con la crítica que Schechtman presentó al modelo del Yo de Strawson, señalando que los Yoes son fluidos y amorfos. Se trata del problema de cómo un Yo sustituye a otro, o en otros términos, de qué características tiene la identidad de los Yoes. En este punto coincido con Schechtman en que el modelo del collar de perlas propuesto por Strawson no se corresponde a la experiencia fenomenológica de los

13 A este respecto recuérdese lo dicho en la nota al pie número 9 de este mismo capítulo.

Yoes, pues a menudo es difícil decir si uno continúa siendo o no el mismo Yo que era tiempo atrás. Ante esta evidencia no podemos hacer más que reconocer que la identidad de los Yoes es indeterminada, y que en ocasiones no será posible establecer de forma categórica donde acaba un Yo y empieza otro. Aunque es posible que la transición entre dos Yoes sucesivos se dé de forma radical e instantánea –quizá a causa de un suceso traumático– normalmente las transiciones de uno a otro están más diluidas en el tiempo, existiendo periodos en los que el Yo que existe en un cuerpo no es ni idéntico ni diferente a los inmediatamente anterior y posterior.

5.3. El mundo de las personas

Una vez hemos referido brevemente los mundos en los que existen y las características que tienen los seres humanos y los Yoes, ha llegado el momento de hacer lo propio con las personas. Como se recordará, ya en el tercer capítulo avancé una definición preliminar de lo que debíamos entender por «persona», haciendo hincapié en su dimensión social. En este sentido, el mundo en el que deben definirse creo que es un mundo descrito en términos sociales que requiere de una actitud interpretativa distinta a las que vimos en el capítulo anterior. Por este motivo, antes de dar una definición de «persona» será necesario aclarar en qué consiste esta actitud interpretativa, y de la que al contrario de lo sucedido en los dos casos anteriores, no contamos con una analogía basada en el juego de la vida. Para exponerla me inspiraré además de en la forma en que Dennett define la actitud intencional, en la teoría de la construcción de la realidad social de John Searle (1995).¹⁴

5.3.1. La actitud colectiva

Imaginemos lo siguiente. Tenemos una explanada de césped de forma más o menos rectangular. En dos extremos opuestos del rectángulo hay, en cada lado, dos montones de ropa colocados aproximadamente a la misma

14 Seguramente Searle se sentiría indignado de saber que utilizo su teoría para apuntalar una parte de mi visión anti-realista, pues su postura ontológica es radicalmente opuesta a la mía. Como muestra, un botón. Las dos primeras frases de la introducción de *La construcción de la realidad social* dicen así: «Vivimos exactamente en un mundo, no en dos, en tres o en diecisiete. Hasta donde sabemos, los rasgos más fundamentales de ese mundo están descritos por la física, la química y el resto de ciencias naturales» (Searle, 1995, p. 19).

distancia entre sí. Dentro del rectángulo hay veintidós individuos vestidos con ropa de deporte, aunque no parece haber ninguna distinción especial entre ellos. Todos parecen interesarse por un balón al que mueven dándole patadas. De repente uno de los individuos golpea el balón y éste pasa entre los montones de ropa de uno de los extremos. Diez de los hombres corren hacia el jugador que ha golpeado el balón para abrazarlo. Los otros once hombres parecen tristes. Sorprendentemente todo el mundo parece calmarse tras ello. Ya no corren y han dejado de preocuparse por el balón. Entonces uno de los hombres tristes lo agarra con la mano y con calma se dirige hasta la mitad del rectángulo. A un lado se sitúan los hombres que estaban tristes y al otro lado los que estaban felices. Sin mediar palabra el hombre que había agarrado la pelota le da una patada y todos comienzan a correr de nuevo tras ella.

Seguramente todos los lectores de estas páginas saben que lo que acabo de describir es un gol en un partido de fútbol *amateur*. Sin embargo, ¿por qué sabemos que se trata de un partido de fútbol? Sin duda la situación no parece implicarlo de un modo evidente. Podemos imaginar a alguien que no sepa de la existencia del fútbol ni de ningún otro deporte en equipo. Si este sujeto viese la situación anterior, aparentemente vería todos los hechos que a nosotros nos permiten saber que se trata de un partido de fútbol. Sin embargo nuestro hipotético observador no podría deducir lo mismo que nosotros. De hecho posiblemente no fuese capaz de detectar jugadas concretas en el campo, ni siquiera dos equipos diferenciados, por lo que tendría la sensación de que los jugadores se mueven más o menos al azar sin ningún objetivo definido. Aunque a nosotros nos parezcan evidentes los motivos, que todos los jugadores se detuviesen de repente al meter un gol para poco después volver a empezar sería para él un acto misterioso.

Cambemos un poco la situación. Pongamos que nuestro partido de fútbol ya no es tan *amateur*, y que los jugadores llevan uniformes bien diferenciados. Además los montones de ropa de los extremos se han transformado en porterías y las líneas del campo están perfectamente pintadas. Aunque nuestro hipotético observador seguiría sin saber que está viendo un partido de fútbol, no tardaría en comprender que los jugadores no se mueven al azar y podría empezar a interpretar lo que está viendo. En seguida se daría cuenta de que los que llevan la camiseta de color blanco tienen un objetivo en común y *colaboran* para conseguirlo: *quieren* meter el balón en una de las porterías y evitar que los de la camiseta azul lo metan en la otra. El que los elementos del juego estén más claramente representados le facilitaría a nuestro observador hacer una interpretación de lo que está viendo porque los patrones de comportamiento de los

jugadores serían más evidentes. Aunque seguiría viendo veintidós hombres en el campo, se daría cuenta de que no importan tanto sus movimientos individuales como los movimientos de cada equipo en su conjunto. Pese a la multiplicidad de jugadores, o quizá debido a ello, le sería más fácil interpretar el juego una vez que se diese cuenta de que todos ellos se comportan como si en realidad fuesen únicamente dos entidades. Una vez dado este paso, la interpretación le resultaría mucho más sencilla. Sólo tendría que aplicar algo parecido a la actitud intencional de la que hablamos en el capítulo anterior, y suponer que esas dos entidades son un tipo especial de agentes racionales, con sus propias creencias y deseos, y que por tanto actúan del mejor modo que les es posible para conseguir sus objetivos. En otras palabras, sólo tendría que suponer que existe la intencionalidad colectiva.

Tras esto nuestro observador también podría darse cuenta de que además de la intencionalidad colectiva de cada equipo por separado, existen una serie de creencias que ambos grupos comparten. Cuando el equipo blanco mete gol, el equipo azul no se muestra indiferente, sino que reconoce que se trata de una parte del juego que les es adversa. Del mismo modo, los jugadores del equipo blanco no se lanzan a por el balón inmediatamente para poder meter otro gol, sino que esperan pacientemente a que el equipo azul saque, pues reconocen que así es como el juego funciona. Es decir, los dos equipos creen que existen unas reglas que deben cumplir y aceptan seguirlas, pues sólo así cobra sentido su actuación: se trata de un partido de fútbol únicamente porque los dos equipos creen que están jugando al fútbol, porque creen que están haciendo algo conjuntamente. Si uno de ellos no compartiese esa creencia ya no tendríamos un partido de fútbol, sino otra cosa distinta. Por poner un ejemplo todavía más claro, dos hombres están boxeando sólo si ambos piensan que eso es lo que están haciendo y el público que les rodea reconoce que está viendo un combate. Si sólo uno de los combatientes lo creyese lo que tendríamos sería una agresión. Si por el contrario fuese el público el que no reconociese dicha actividad, lo que tendríamos sería una pelea que la policía debería detener para evitar que dañen a terceros.

En este sentido podemos ver que existen creencias colectivas que son compartidas no sólo por unos cuantos individuos, sino por una sociedad en su conjunto. Y que éstas son de lo más variado. La gente efectivamente no llama a la policía para detener un combate de boxeo porque todos creen que los combatientes están realizando una actividad conjunta. Pero igualmente, aunque la mayoría de la gente no salga a la calle vestida con ropa estrafalaria, nadie se ríe de los policías o de los bomberos por hacerlo, pues creen que no se trata de un disfraz, sino de un uniforme. Y si

nadie se escandaliza y nos toma por locos cuando parece que vamos hablando solos con unos aparatos rectangulares pegados a la oreja es porque creen que esos aparatos son teléfonos y que por tanto hablamos con otra persona que está ausente, aunque nadie se moleste en comprobarlo. Podemos ver así que estas creencias colectivas pueden referirse no sólo a actividades, sino también a sujetos u objetos concretos.

Pero lo más interesante de este tipo de creencias colectivas es que no sólo «se refieren a» sujetos u objetos, sino que en cierto sentido parecen ser constitutivas de los sujetos u objetos a los que se refieren. Una persona no es un policía porque crea serlo, o por llevar un tipo de uniforme, sino porque todo el mundo acepta su autoridad y su papel como garante de la seguridad ciudadana. Si no fuese así, sólo sería un loco disfrazado que intenta imponer su voluntad por la fuerza. En este sentido creo que basándonos en este tipo de creencias colectivas podemos definir una actitud interpretativa diferente a las que vimos en el capítulo anterior y que nos permitiría describir una nueva serie de entidades. Adoptar esta actitud, a la que podemos llamar *actitud colectiva*, consistiría en suponer que un conjunto de sujetos poseen una serie de creencias, deseos e intenciones comunes, y tratar de definirles tanto a ellos mismos como al entorno que les rodea en función de estos patrones de intencionalidad colectiva que les hemos adscrito.

Adoptar esta estrategia interpretativa nos permite detectar una serie de patrones en la REALIDAD que no son accesibles de otro modo, como los que hemos visto que están implicados en un partido de fútbol o en el reconocimiento de la autoridad de un policía. Pero al mismo tiempo, y al igual que sucede con la actitud intencional, podemos adoptar la actitud colectiva para interpretar cualquier aspecto del mundo, aunque en unas ocasiones parecerá más pertinente que en otras. Interpretar que algo es una silla porque todo el mundo cree que es una silla no parece de mucha utilidad. Sin embargo interpretar que algo es un combate de boxeo porque todo el mundo cree que es un combate de boxeo impide que lo confundamos con una pelea en la que deba intervenir la policía. Del mismo modo la creencia colectiva de que algo es dinero explica por qué cuando vamos a hacer la compra el dependiente no nos pregunte con suspicacia si el valor de los billetes que le estamos dando es realmente el que decimos que tienen. Sabe que todo el mundo da por sentado el valor del dinero, y que por tanto no podría darse el caso de que si en el futuro él quisiese utilizar los billetes que le hemos dado le dijese que en realidad sólo valen la mitad. En este sentido, y en base a lo que explicamos en el capítulo anterior, podemos construir interpretaciones amplias de la

REALIDAD que presupongan la existencia de mundos reales mediante la detección de patrones de intencionalidad colectiva.

Un aspecto que me gustaría aclarar antes de seguir adelante es el referido a la aparente labor constitutiva de las creencias colectivas en lo referido a las entidades definidas desde esta actitud. Desde el punto de vista de Searle, la realidad social se constituye realmente gracias a las creencias colectivas, de tal modo que si no hubiese creencias colectivas no habría entidades sociales. Sin embargo esto genera un problema, y es el de quién es el sujeto de dichas creencias y si realmente las tiene. Una característica sorprendente de las creencias colectivas es que no tienen por qué ser compartidas conscientemente por ninguno de los individuos de la sociedad a la que adscribimos una creencia. Por poner un ejemplo, nadie tiene por qué creer conscientemente que en las sentencias castellanas se pueden introducir adverbios para que de hecho se introduzcan. Aunque nadie supiese lo que es un adverbio ni supiese distinguir las funciones lingüísticas que cumple, los hablantes del castellano seguirían utilizando adverbios, por lo que estaríamos justificados para atribuirles la creencia colectiva de que en castellano se pueden utilizar adverbios. En este sentido, al igual que sucede con otras actitudes interpretativas, estamos justificados para utilizar la actitud colectiva siempre que el atribuir creencias colectivas a una sociedad redunde en un mayor poder explicativo, independientemente de si REALMENTE existe un sujeto que tenga dichas creencias. La REALIDAD es ajena a nuestros intentos de descripciones de la misma y en este sentido las creencias colectivas que atribuimos a una sociedad para explicar su comportamiento ni son REALES ni constituyen nada REAL, sino que simplemente describen patrones sociales observables. Por otro lado, y como ya señalamos en el capítulo anterior, al intentar describir la REALIDAD postulamos la existencia de una realidad externa a nosotros que no existiría si no la hubiésemos descrito. En este sentido las creencias colectivas sí son constitutivas de realidad, pues como ya dijimos las descripciones de mundos en cierta manera construyen mundos.

Otro aspecto que convendría hacer explícito son los motivos por los que considero que la intencionalidad colectiva es un fenómeno distinto a un conjunto de intencionalidades individuales y que por tanto los mundos descritos desde la actitud colectiva son esencialmente irreducibles a los mundos descritos desde la actitud intencional. En primer lugar, ya he señalado que las creencias colectivas no tienen por qué ser compartidas conscientemente por ninguno de los miembros de la colectividad que posee una creencia, aunque podría argumentarse que dichos miembros pueden poseerla sin ser conscientes de ello. Pero en segundo lugar, si

intentamos comprender los fenómenos de intencionalidad colectiva reduciéndolos a fenómenos de intencionalidad individual nos encontramos con un problema de regresión infinita. Si queremos reducir una creencia colectiva simple como «nosotros creemos X» al lenguaje de la intencionalidad individual nos encontramos con que debería traducirse por la suma de «yo creo X», «tu crees X», «yo creo que tu crees X», «tu crees que yo creo X», «yo creo que tu crees que yo creo X», «tu crees que yo creo que tu crees X», etc. Evidentemente este esquema tendría que seguir reproduciéndose hasta el infinito, lo que creo que es motivo suficiente para renunciar a encontrar una respuesta reduccionista (Searle, 1995, pp. 41-44). Si los filósofos normalmente tratan de hacer este tipo de reducción es porque creen que si consideramos la intencionalidad colectiva como un fenómeno primitivo tenemos que comprometernos con la existencia de algo parecido a un espíritu del mundo que tenga las creencias de las que estamos hablando. Sin embargo, como acabo de señalar y al igual que sucede con todas nuestras interpretaciones del mundo, la actitud colectiva no nos obliga a comprometernos con la existencia REAL de ninguna de las cosas que postula, ni tiene tampoco la obligación de ofrecer una explicación satisfactoria de todos los fenómenos que existen. Adoptar la actitud colectiva tiene la consecuencia desagradable de que no es claro quién es el sujeto de las creencias colectivas que postulamos (*cf.* Carr, 1986, pp. 128-130), al igual que también las otras actitudes interpretativas que vimos en el capítulo anterior tienen sus propias limitaciones.

Una última objeción que se le puede poner a esta manera de entender la realidad social es que tal y como la estamos describiendo parece implicar que todos los objetos que se definan desde la actitud colectiva estarán definidos de forma circular. Si adoptamos esta actitud parece que estamos comprometidos a decir que alguien es un policía porque creemos que es un policía. Que algo es dinero porque creemos que es dinero. Y que algo es un combate de boxeo o un partido de fútbol porque creemos que es un combate de boxeo o un partido de fútbol. Este tipo de definiciones circulares no parecen ser informativas, por lo que si efectivamente estuviésemos comprometidos a definir de este modo todas las entidades que componen los mundos que podemos describir desde la actitud colectiva lo que estaríamos describiendo serían mundos ininteligibles. Sin embargo eliminar esta circularidad es sencillo. Basta con especificar por qué se cree que algo es un policía, dinero, un combate de boxeo o un partido de fútbol. Esto es, si decimos que algo es dinero porque creemos que es dinero, lo que estamos queriendo decir es que creemos que puede utilizarse en una serie de prácticas como la compra de productos, el pago de servicios, el saldo de deudas, etc. Para

eliminar la circularidad en las definiciones basta con explicitar los motivos de las creencias poniéndolos en relación con las prácticas en las que intervienen (*vid.* Searle, 1995, pp. 68-70).

5.3.2. *La tesis social de la Narratividad*

Una vez que hemos descrito la actitud colectiva podemos decir que el mundo en el que existen las personas es el mundo que obtenemos si aplicamos esta actitud a la interpretación de la sociedad occidental actual. Y que por tanto las personas no son más que aquello que nosotros creemos que son las personas. Curiosamente Dennett ya sugirió una definición en este sentido antes de empezar a interesarse por el tema del Yo narrativo. Sin embargo esta idea le hacía pensar que, de ser cierta, debería redundar en una disminución de la importancia que atribuimos al concepto:

Podría resultar, por ejemplo, que el concepto de persona no sea sino un nebuloso término honorífico que gustosamente aplicamos a nosotros mismos y a los demás, como mejor nos parece, guiados por nuestras emociones, nuestra sensibilidad estética, cuestiones de costumbre y similares –así como los *chic* son sólo aquellos que pueden lograr que los consideren *chic* otros que a sí mismos se tienen por *chic*. El ser una persona es, sin duda, *algo* similar a esto, y si no fuera más que esto, tendríamos que reconsiderar, de ser posible, la importancia que hoy atribuimos al concepto. (Dennett, 1976, p. 6)

A diferencia de Dennett, no considero que el hecho de que ser personas sea debido a que nosotros mismos nos consideramos así le reste importancia al concepto. Sin embargo, y como acabamos de señalar, una definición de este tipo no es informativa, por lo que para terminar de definir qué es lo que considero que son las personas convendrá hacer explícito cómo se forman las creencias colectivas acerca de ellas que nos hacen adscribirnos a nosotros mismos esa categoría.

En primer lugar, merece la pena señalar cómo encajan los conceptos de «persona» que hasta ahora han manejado en el NPN Schechtman y Zahavi con el que estoy tratando de definir. Como se recordará, ambos autores defienden que la persona es una entidad construida narrativamente que debe poder dar cuenta de la dimensión social de nuestra existencia. Sin embargo aún consideran que las personas se definen desde el nivel intencional. Para ellos lo relevante es la narración que cada persona hace de sí misma, pues ésta es lo que la constituye. Esta actitud desde mi punto de vista es poco productiva, pues por un lado resulta difícil distinguir entre la persona y el Yo si ambos se constituyen mediante una narración en

primera persona; y por otro, creo que no reflexiona lo suficiente sobre lo que significa poder dar cuenta de nuestra existencia social. La sentencia de MacIntyre que referí en el segundo capítulo sobre la importancia de deber rendir cuentas de todo lo que los demás nos exijan me parece algo que queda totalmente fuera del rango de lo que una narración en primera persona como la que proponen Schechtman y Zahavi puede ofrecer.¹⁵ En este sentido creo que puede ser útil revisar los argumentos de los primeros teóricos del PNC que dieron lugar a la tesis ética de la Narratividad, pues como señalé en aquel momento su objetivo era dar cabida a la dimensión social de nuestra existencia, sobre todo en lo referido a la necesidad de que las personas respondan a la imputación de identidad estricta, algo necesario para hablar de responsabilidad moral.

Especialmente interesante a este respecto es la postura de MacIntyre, quien como se recordará señalaba el hecho de que no sólo somos un personaje en nuestras propias narraciones, sino también en las narraciones de los que nos rodean. Desde el mismo momento en que nacemos, o incluso antes, nuestros padres comienzan a contar historias sobre nosotros. Historias que necesariamente nos afectan, pues nos resulta imposible contar nuestra propia historia sin tener como referencia las historias de los demás.¹⁶ De este modo sucede que nuestra propia narración se encaja en las narraciones de otros resultando que en definitiva no somos más que meros coautores de nuestra historia (MacIntyre, 1981, p. 263). Ricoeur parece darle la razón a MacIntyre sobre este aspecto (Ricoeur, 1990, pp. 161-162), y por su parte no creo que Taylor tuviese problemas en aceptar algo parecido, dado su especial interés por vincular su teoría de la identidad narrativa con los problemas éticos que aparecen en contextos sociales. Pero no sólo ellos, sino que también Dennett parece proclive a comulgar con la idea:

«Llámenme Dan» oyen salir de mis labios, y ustedes obedecen, pero no llamando Dan a mis labios, o a mi cuerpo, sino llamándome Dan a mí, la ficción de un teórico creada por... bueno, no por mí, sino por mi cerebro, actuando en concierto durante años con mis padres, hermanos y amigos. (Dennett, 1991a, p. 439)

15 «Soy para siempre cualquier cosa que haya sido en cualquier momento para los demás, y puedo en cualquier momento ser llamado para responder de ello, no importando cuánto pueda haber cambiado» (MacIntyre, 1981, p. 267)

16 En este sentido también puede ser pertinente citar a Todorov, para quien el Yo actual es una escena en la que está presente, entre otros, un Yo reflexivo que es «imagen de la imagen que los demás tienen de nosotros –o más bien de aquella que imaginamos estará presente en sus mentes–» (Todorov, 1995, pp. 25-26).

De este modo podemos ver que para algunos de los primeros teóricos del PNC nuestra identidad consistía por un lado en nuestro relato subjetivo sobre nuestra propia vida, pero también y al mismo tiempo en el relato que hacen sobre nosotros los que nos rodean. El primero nos permitía acercarnos al modo en que nuestra conciencia percibe subjetivamente el mundo, solucionando así el primero de los problemas que achacaban al planteamiento objetivo. Por su parte los segundos servirían de contrapunto para conseguir que nuestro relato no se desviase del marco social en el que nos desarrollamos, atendiendo al otro fallo que veían en el PO.¹⁷ En el fondo de esta propuesta latía la presuposición de que uno y otros eran distintas caras de un mismo relato que podían ser reunidas y consideradas como una única narración en cuyo centro nos encontraríamos inequívocamente nosotros. Pero si atendemos a nuestra propia vida no resulta difícil encontrar ejemplos de situaciones en las que nuestro relato sobre un determinado evento difiere del relato que cuentan nuestros amigos. Normalmente se trata de divergencias sin importancia –¿ocurrió un sábado o un domingo?–, por lo que una revisión superficial del caso nos permite llegar a un acuerdo, o descartar las diferencias por irrelevantes. Pero dado que nuestra memoria no es fiable cuando se trata de reconstruir eventos concretos siempre existe la posibilidad de que existan versiones incompatibles que difieran en aspectos sustanciales. Cuando éste es el caso, normalmente parece que la visión de la mayoría se impone, como se pone de manifiesto si pensamos en un juicio por asesinato en el que el acusado niega haber estado en la escena del crimen pero una gran cantidad de testigos lo sitúa allí. Esto sin duda sería un problema para los teóricos del PNC, que necesitaban que la narración postulada atendiese al mismo tiempo a las dimensiones subjetiva y social de nuestra existencia. Sin embargo en el marco del NPN lo que importa cuando tratamos de personas son las creencias colectivas que las constituyen. En este sentido que sean los relatos de los demás los que en principio determinen nuestra identidad no parece un problema.

Teniendo en cuenta estas reflexiones podemos formular lo que llamaré la *tesis social de la Narratividad*, y que dice que las creencias colectivas que constituyen la identidad de las personas se forman mediante una negociación entre nuestro propio relato de vida y los relatos que cuentan los demás sobre nosotros. Sin embargo, como acabo de señalar, se trata de una negociación a primera vista asimétrica, en donde se hallan en tensión

17 Podemos ver que esta presencia de los relatos de otros se encuentra mucho más difuminada por ejemplo en la teoría de Schechtman, que sustituye a los demás como mecanismo de control por lo que denomina la «reality constraint» (Schechtman, 1996, pp. 119-130).

dos polos: la narración propia y las narraciones de los otros. Un aspecto que pese a haber sido en general ignorado en el ámbito anglosajón, curiosamente ha sido objeto de estudio de dos filósofos españoles, Manuel Cruz y Fernando Broncano.

El motivo de que esta tensión haya sido ignorada en el ámbito anglosajón es que la postura más extendida entre los filósofos analíticos es que la identidad es decidible sólo desde la primera persona. Lo que importa cuando nos preguntamos por la identidad de alguien es lo que la propia narración del sujeto señale. Si alguien asegura que ya no es el mismo Yo que era en el pasado (entiéndase que el término «Yo» aquí es equivalente al término «persona»), es que en su cuerpo han habitado al menos dos Yoes sucesivos. Pero tal y como señala Cruz, se trata de una presuposición cuanto menos discutible:

La cuestión fundamental, a nuestro modo de ver, es la siguiente: ¿constituye el testimonio de ese individuo, *extrañado* ante su yo pasado, la prueba irrefutable de que realmente no tenemos derecho a hablar de la misma persona? Irrefutable, por lo pronto, desde luego que no. Porque, para que lo fuera, esta pluralidad de yoes debiera poder ser reivindicada en cualesquiera otros contextos. (Cruz, 1999, p. 38)

Aceptar esta postura supone dar pie a lo que Broncano denomina el problema del egocentrismo. Si quien decide es el propio sujeto, las narraciones de los otros no son más que ruido en la ecuación, algo de lo que se puede prescindir (Broncano, 2013, p. 188). Pero construirse una identidad al margen de los otros es una quimera, o en el mejor de los casos un camino hacia la soledad más absoluta, pues tener identidad no es otra cosa que una forma de ser aceptados por los demás (Cruz, 2005, pp. 81-82). Sin embargo la solución no consiste en irse al otro extremo, como podría deducirse de lo que decíamos unos párrafos atrás. Considerar la identidad como el producto únicamente de las narraciones que hacen los otros tampoco serviría para hacer justicia a nuestra existencia social. Aunque hacerlo nos permitiese justificar que las personas puedan responder a la imputación de identidad estricta. En palabras de Cruz:

No basta con disponer de un hipotético agente a quien endosarle la autoría de lo ocurrido: eso equivaldría a convertir el discurso ético en un ciego mecanismo de atribución de responsabilidades. Se precisa asimismo que haya alguien dispuesto a reclamar ese acto como suyo, a hacerse cargo de él. (Cruz, 2005, p. 52)

De este modo para definir a las personas nos vemos en la obligación de encontrar un equilibrio entre ambos polos. La *negociación narrativa* que

da lugar a nuestra identidad tiene que partir del hecho de que ninguna de las dos posturas puede imponerse arbitrariamente a la otra. Por tanto las creencias colectivas que dan lugar a dicha negociación tienen que contener la idea de que cada persona tiene una cierta *autonomía*, al mismo tiempo que también *depende* del resto de componentes del espacio social. «Autonomía que no puede eliminar la dependencia sin convertir al sujeto en un átomo autista; y dependencia que no puede ahogar la trayectoria del sujeto en un cúmulo de chantajes efectivos» (Broncano, 2013, p. 104). En una definición alternativa podemos decir entonces que ser persona «es moverse con adecuada habilidad entre dos fuentes de autoridad: la *autoridad propia*, el autogobierno o autonomía, y la *autoridad del mundo*, en el que cobran una especial relevancia los otros y la autoridad que implica su presencia» (Broncano, 2013, p. 69).

Considerar que las personas se constituyen de este modo evidentemente implica una serie de requisitos previos. Entre ellos algunos que ya señalamos al desarrollar el experimento mental del jefe mafioso. Por un lado, para que haya una negociación narrativa es necesario que las personas tengan una serie de capacidades mentales complejas: racionalidad, capacidad de comprenderse a sí mismas y a los demás de una forma narrativa, capacidad de comunicar dicha comprensión mediante un lenguaje articulado y autoconciencia. Por otro, para que la negociación tenga sentido es necesario que las personas también sean capaces de considerarse a sí mismas y a los demás como personas, lo que implica considerarlas como agentes autónomos con una serie de derechos y responsabilidades (*cf.* Dennett, 1976).

De este modo podemos ver que la creencia de que somos personas queda lejos de ser un simple título honorífico como sugería Dennett. Por un lado, se trata de una creencia fundada en una serie de factores que hacen de las personas unas entidades cuyo comportamiento es especialmente complejo. Por otro, se trata de una creencia que en cierto sentido, y como ya señalamos, crea realidad. Las personas no son autónomas simplemente en virtud de sus cualidades intrínsecas, sino porque se les reconoce dicha autonomía. No quedan lejos los tiempos en los que en España las mujeres necesitaban el permiso de su padre o de su marido para abrir una cuenta bancaria, lo que en última instancia no era más que una forma de decir que no eran personas. En este sentido debemos tener en cuenta que precisamente porque se nos reconoce como personas podemos participar con normalidad en la vida social.¹⁸

18 A este respecto es interesante leer lo que Amélie Rorty señala sobre la doble raíz del concepto de «persona». Por un lado tiene que ver con la analogía que se establece entre los personajes teatrales (*dramatis personae*) y los roles

El último aspecto entonces que queda por tratar referido a las personas es el de su identidad y sus condiciones de supervivencia. Ya en el tercer capítulo dije algo al respecto, y en el siguiente apartado profundizaré en ello. Por ahora lo único que me gustaría señalar es que, a diferencia de los Yoes y los seres humanos, la identidad de las personas es estricta. Aunque en principio una negociación narrativa podría terminar con la conclusión de que una persona actual sólo es en parte la misma que otra persona pasada, la sociedad en la que vivimos nos fuerza a tomar decisiones. Resulta difícil imaginar un juicio en el que el juez declare al acusado «ni inocente ni culpable» porque al mismo tiempo sea y no sea la persona que cometió el crimen. Necesitamos que haya responsables de las acciones y actuamos en consecuencia llevando las negociaciones hasta el punto en el que se toma una decisión. En este sentido debemos considerar que la propia identidad personal no es una opción que podamos tomar o desechar. No se trata de un elección entre aceptar el juego de la responsabilidad moral o vivir al margen. Entre aceptar lo que los demás demandan de nosotros o tratar de imponer nuestra independencia. Existir como persona es existir en sociedad, y si queremos existir debemos por tanto entrar en la negociación:

La identidad es la entidad que nos atribuyen los otros. No es una opción, sino un destino. Un destino que, es cierto, en ocasiones se torna carga, pero del que no podemos abdicar. Este carácter irrenunciable [...] no es un *desideratum* ético sino un imperativo de supervivencia. (Cruz, 2005, pp. 83-84)

Resulta interesante leer desde esta perspectiva por ejemplo *Uno, ninguno y cien mil* de Luigi Pirandello. El protagonista de la novela, Vitangelo Moscarda, se siente turbado cuando descubre que los demás

sociales que desempeñamos. Por otro procede del ámbito legal y se relaciona con la necesidad de atribuir responsabilidades. Aunque ambas raíces son independientes, pues todos tenemos algún rol social pero no necesariamente somos legalmente responsables de nuestras acciones, en realidad cuando a alguien se le niega el estatuto legal de «persona» se le impide también su desarrollo social. La teología cristiana –al contrario que la jurisprudencia romana– necesitaba que todo ser humano alcanzase la condición de «persona» (esto es, que se le reconociese la libertad para actuar según su voluntad) para poder ser juzgado por Dios, abogando así por una fusión de ambas raíces en un único concepto que es el que se hereda en la discusión sobre la identidad personal (Rorty, 1976, pp. 309-310). Sin embargo, y como veremos más adelante, esta fusión cristiana es también una fuente de problemas para el tema que nos ocupa, pues provoca que se confundan los conceptos de «persona» y de «ser humano».

tienen una imagen de él muy distinta a la suya propia. Ante esta situación se obsesiona con afirmar su propia auto-imagen, su narración de sí mismo, intentando que todos abandonen las antiguas ideas que tienen sobre él. El resultado, como no podía ser de otra manera, es la locura. La única forma que tiene de que todos tengan la misma imagen de él es proyectar la imagen de un loco, es decir, la imagen de alguien que ya no cumple los requisitos para ser una persona pues ha perdido la razón. De este modo, al negarse a negociar su identidad, la sociedad negocia en su ausencia, llegando a la conclusión de que al ser un sujeto irracional la narración que tiene de sí mismo es irrelevante. Vitangelo queda inhabilitado como persona, si bien no legalmente, sí en la práctica, pues el único modo que tiene de evitar el juicio mediante el que sus allegados quieren declararle incapaz es renunciar a todas sus posesiones y retirarse a vivir en un hospicio en las afueras de la ciudad. El caso del protagonista de la novela ejemplifica así este carácter irrenunciable de la identidad personal. Sólo aceptando participar en la negociación narrativa que nos constituye podemos sobrevivir como personas.

5.4. Resolviendo los experimentos mentales

En el tercer capítulo presenté el experimento mental del jefe mafioso para tratar de demostrar que cuando nos preocupamos por nuestra supervivencia nos deberían importar en la misma medida las tres entidades que se distinguen en el NPN. Sin embargo aquel experimento tenía sus limitaciones, especialmente porque aún no contábamos con unos conceptos desarrollados de «ser humano», «Yo» y «persona». Por este motivo mi intención en aquel momento no era analizar en profundidad en qué consistía la identidad de cada una de las entidades y hasta qué punto podíamos sentirnos identificados con ellas, sino tratar de mostrar del modo más sintético posible que sus condiciones de supervivencia eran independientes y que podíamos preocuparnos por nuestro futuro en tanto cualquiera de las tres. Una vez contamos con unas definiciones suficientemente desarrolladas espero poder hacer más claros estos aspectos. Para ello me gustaría volver sobre los experimentos mentales que provocaron el abandono del planteamiento objetivo —especialmente los que pusieron de relieve los problemas de la duplicación y de la fusión— para mostrar cómo desde el nuevo planteamiento narrativo éstos tienen solución.

Pensemos en primer lugar en los casos de duplicación, como el de los dobles de Guy Fawkes, el de los trasplantes de medio cerebro o el del

teletransporte. Como se recordará, estos experimentos presentaban situaciones en las que, ya fuese siguiendo un criterio físico –en el caso de los trasplantes– o un criterio psicológico –en los casos de Guy Fawkes y del teletransporte–, al final nos encontrábamos con que teníamos idénticos motivos para atribuir la identidad de la persona original a dos entidades distintas, por lo que debíamos admitir que la identidad era una relación indeterminada. Por su parte los partidarios del PNC podrían interpretar estos casos con mayor fortuna que los del PO, pues entendiendo a las personas como entidades mentales encarnadas en cerebros concretos podrían decir que ni los dobles de Guy Fawkes ni las réplicas creadas en el caso del teletransporte podrían aspirar a la identidad de la persona original, pues tendrían una base cerebral distinta (*vid.* Slors, 1998). Sin embargo no está tan claro que pudiesen dar una respuesta adecuada al caso de los trasplantes de medio cerebro. Dennett sugiere que desde su postura se podría argumentar que medio cerebro no sería suficiente para dar soporte a un Yo, aunque es consciente de que siempre se puede ir estrechando los límites, de modo que en algún momento sería posible que no estuviésemos seguros de si una base cerebral concreta podría albergar un Yo o sólo un cuasiYo, por lo que acepta también la conclusión parfitiana de que la identidad es indeterminada (Dennett, 1991a, pp. 429-436).

Desde el nuevo planteamiento narrativo la situación es diferente. Como hemos dicho, el NPN distingue entre tres entidades, la identidad de dos de las cuales –seres humanos y Yoes– es indeterminada. Sin embargo sostiene también que la identidad de las personas es determinada, estricta. Bajo este punto de vista supone una mejora con respecto al PO y al PNC, pues nos permite mantener nuestra intuición de que siempre debe haber una respuesta a la pregunta «¿seguiré vivo mañana?». Enfrentados entonces a los experimentos mentales de duplicación, desde el NPN podemos interpretarlos de un modo más acorde con ésta creencia. Con respecto a los seres humanos y los Yoes podemos decir que ni los dobles de Guy Fawkes ni las réplicas creadas por una máquina teletransportadora serían los mismos que los seres humanos o Yoes originales, mientras que en el caso del trasplante de medio cerebro –que podemos interpretar como un caso de trasplante de medio cuerpo para hacer más clara la situación con respecto al ser humano– lo que tendríamos sería un caso que evidenciaría que su identidad es indeterminada. Sin embargo, en lo que respecta a las personas, la situación adopta una forma más compleja, pues la respuesta a cada caso dependerá de factores que no estaban incluidos en los experimentos tal y como los formulé en el primer capítulo.

En el caso de los dobles de Guy Fawkes, hemos de suponer que el Guy Fawkes original antes de morir no contemplaba la posibilidad de

sobrevivir como persona gracias a una réplica suya. Al mismo tiempo la sociedad actual en la que presumiblemente se producirían de forma espontánea sus dobles no creo que estuviese muy dispuesta a aceptar la reencarnación de un hombre al que nadie conoció en vida y del que apenas se conocen unos escasos datos sobre su intento de atentado contra el parlamento británico. En este sentido la negociación narrativa que daba vida a la persona Guy Fawkes se interrumpió con su muerte en enero de 1606. Ni él ni sus contemporáneos compartían la creencia colectiva de que podría sobrevivir, ni tampoco nosotros la tenemos. En consecuencia ninguno de los dobles de Guy Fawkes podría reclamar la identidad de la persona original porque no existen las creencias colectivas que pudiesen dar lugar a una negociación narrativa en ese sentido.

Por su parte en el caso del teletransporte la situación es diferente. Tal y como señalábamos, Douglas Quaid vive en un mundo futuro en el que los viajes mediante teletransportación son habituales. La mañana antes de que Quaid utilizase el teletransporte por primera vez, su propia esposa –a la que él sigue considerando como su esposa– le señaló que ella se había teletransportado muchas veces y seguía viva. Y el propio Quaid piensa que va a sobrevivir, pues finalmente pulsa el botón que inicia la teletransportación con la creencia de que sobrevivirá gracias a una réplica suya creada en Marte. En este sentido podemos ver que existen en esa sociedad futura las creencias colectivas necesarias para dar lugar a una negociación narrativa que indique que Douglas Quaid sobrevivirá como persona gracias a su réplica. El caso un poco más complejo en el que la máquina de teletransporte no desintegra el cuerpo original de Quaid no alteraría el resultado. Su cuerpo original mantendría su identidad como ser humano y como Yo, pero su identidad como persona habría sido teletransportada a Marte como en el caso del teletransporte simple. Sería el Quaid marciano el que seguiría con su vida como persona, yendo a su trabajo y acostándose con su mujer, pues eso es lo que tanto él como la sociedad en la que vive esperarían. Si el Quaid terrestre quisiese ir al banco para sacar al dinero de Quaid, le debería ser denegado el acceso. Si quisiese intimar con su esposa, ésta debería rechazarlo. Aunque mantuviese las identidades del ser humano y el Yo Quaid originales, esto no le daría permiso para apropiarse de la identidad social de la persona Quaid. De este modo podemos ver que en el caso del teletransporte, y a diferencia del caso de los dobles de Guy Fawkes, la réplica marciana de Quaid sí sería la misma persona que era el Quaid original.

Finalmente, en cuanto al caso del trasplante de medio cerebro, la situación es de nuevo diferente. Tal y como está descrito el experimento, hemos de suponer que sucede en un futuro cercano en el que nunca se ha

dado el caso de que un doble trasplante de dos mitades cerebrales fuese exitoso. De este modo ni la persona original ni los miembros de la sociedad en la que vive disponen *a priori* de las creencias colectivas necesarias para hacer que la persona sobreviva en uno u en otro cuerpo. Sin embargo, a diferencia del caso de Guy Fawkes, tampoco existen *a priori* creencias que indiquen que la persona ha fallecido. En este caso nos encontramos con que en principio no tenemos una respuesta que indique si la persona ha sobrevivido o ha muerto. La identidad de la persona parecería tan indeterminada como la del Yo o la del ser humano. Sin embargo, como ya hemos señalado, vivimos en una sociedad en la que se tiene la necesidad de no zanjar una negociación narrativa hasta que se llega a un acuerdo sobre la identidad de una persona. En este sentido podemos recordar que, como señalamos en el anexo I, la determinación o indeterminación de la identidad es siempre relativa a un marco conceptual dado (Sanfélix, 1994, p. 258). Cuando nos encontramos con una pregunta sin respuesta siempre es posible fijar nuevos criterios o convenciones y transformar así la identidad de indeterminada en determinada, del mismo modo que los jueces pueden crear jurisprudencia al fallar en casos en los que la ley es interpretable. De este modo no podemos saber qué es lo que pasaría con la identidad personal en el caso del trasplante exitoso de dos mitades cerebrales. Pero de lo que sí podemos estar seguros es de que si se diese el caso, la sociedad tendría la necesidad de adoptar una postura al respecto para conservar la determinación de la identidad.

Si nos fijamos ahora en los casos de fusión de personas comprobaremos que la situación es análoga a la del trasplante de dos mitades cerebrales. El único caso de este tipo que expuse en el primer capítulo era el del científico malvado que decidía, mediante una máquina de su invención, ir sustituyendo las células del cuerpo de una persona por las de otra. De este modo, en una situación intermedia en la que la persona resultante tuviese la mitad de las células de la persona original y la mitad de las células de Napoleón, parecería que la pregunta por su supervivencia sería una pregunta vacía. Sin embargo, aunque podríamos aceptar que su identidad como ser humano o como Yo sería indeterminada, con la identidad de la persona nos encontraríamos de nuevo con un caso ante el que aunque *a priori* no podríamos dar una respuesta, la sociedad haría lo posible para encontrarla *a posteriori*.

A través de estos casos podemos ver que nuestra supervivencia como cualquiera de las tres entidades puede darse de forma independiente. Y al contrario de lo que podría pensarse a raíz del experimento del jefe mafioso, en ello no tiene por qué haber implicado ningún daño para las entidades resultantes. En el caso de Tony Soprano su supervivencia como

ser humano se daba gracias a su cuerpo en coma, lo cual era una situación poco atractiva. Sin embargo podemos ver que sería posible sobrevivir como ser humano en perfectas condiciones físicas aún cuando no se sobreviviese ni como Yo ni como persona. Éste sería el caso por ejemplo si ante una amnesia retrógrada completa el Yo resultante asegurase no ser el mismo Yo que habitaba su cuerpo anteriormente y la sociedad decidiese tomar su testimonio como cierto y ante su nueva forma de comportarse procediese por tanto a negarle los derechos y responsabilidades de la persona que era en el pasado para otorgarle una nueva identidad.¹⁹ Igualmente podríamos detallar casos en los que sería posible sobrevivir como Yo aunque no lo hiciésemos ni como seres humanos ni como personas (por ejemplo mediante un cerebro en una cubeta en una sociedad en la que se considerase que sobrevivir así no equivaldría a sobrevivir como persona), en los que podríamos sobrevivir como personas aunque no lo hiciésemos ni como seres humanos ni como Yoes (esto es lo que sucedería en el caso del teletransporte anteriormente mencionado), o cualquier otra combinación que podamos imaginar.

Otro aspecto sobre el que convendría detenerse aunque sea brevemente es sobre la identificación que podemos sentir hacia las tres entidades y en qué medida podemos alegrarnos por su supervivencia. Ya señalé en el apartado dedicado a los Yoes que la identificación que sentimos hacia ellos es una de las partes más importantes de su definición. Así mismo en el tercer capítulo argumenté a favor de que podemos alegrarnos por nuestra supervivencia como personas. En este sentido ahora sólo me gustaría recordar cómo el experimento del científico malvado, en el que mi cuerpo sería torturado tras haberme sido borrada la memoria, se puede interpretar como un caso en el que primero el científico provocará mi muerte en tanto Yo y posteriormente torturará al ser humano que soy. Dado que el experimento mostraba que teníamos motivos para temer dicha tortura, lo que indirectamente demuestra es que también nos identificamos con el ser humano que somos y que por tanto nos preocupamos por su futuro al menos en la misma medida en que nos preocupamos por nuestro futuro en tanto las otras dos entidades.

Finalmente, gracias a la luz que el NPN arroja podemos entender mejor los motivos por los que el problema de la identidad personal llegaba a conclusiones inaceptables en los marcos del PO y del PNC. Las tres entidades que hemos distinguido poseen sin duda propiedades que

19 Algo así parece suceder en la película *Carretera perdida* (David Lynch, dir., 1997), en la que el protagonista sufre una fuga disociativa, lo cual provoca un cambio en su vida que incluye el ser liberado inmediatamente de la cárcel (*vid.* Muñoz Corcuera, 2010a).

normalmente tratamos de adscribirnos a nosotros mismos. Pero una vez que se analizan dichas propiedades, resulta evidente que no son compatibles entre sí. Sin duda las personas debemos responder a la imputación de identidad estricta que demanda nuestra existencia social, sin embargo se trata de una imputación imposible de hacer si consideramos que únicamente somos una asociación de células o un conjunto de propiedades mentales vinculadas a un cuerpo concreto, pues la identidad de tanto los seres humanos como de los Yoes es una cuestión de grado. En este sentido es comprensible que Parfit llegase a la conclusión de que nuestra identidad era indeterminada y que por tanto no era ésta lo que nos debía importar en nuestra supervivencia. Lo que nos debía importar en nuestra supervivencia (al menos en parte), y lo que hacía que nuestra identidad fuese determinada, era algo que no estaba contemplando: nuestra existencia social. Por otro lado en el marco del PNC las personas se definían únicamente desde la primera persona. Si alguien aseguraba que ya no era la misma persona que en el pasado, no era posible argumentar lo contrario. Sin embargo aquello que somos como seres humanos y como personas no está necesariamente relacionado con lo que subjetivamente creamos ser. Que alguien se vuelva loco por leer demasiados libros de caballerías y crea ser un caballero andante no afecta en absoluto al hecho de que seguirá siendo el mismo ser humano que era anteriormente, ni por supuesto implica que la sociedad en la que vive deba aceptar sus desvaríos. De este modo espero que la distinción entre lo que somos como seres humanos, Yoes y personas sirva para aclarar en qué consiste nuestra identidad a través del tiempo y qué nos debe importar cuando nos preocupamos por nuestra supervivencia o por cuestiones de responsabilidad moral.

CAPÍTULO 6

PERSONAS Y PERSONAJES

En el capítulo anterior traté de desarrollar el NPN de tal modo que fuese posible responder la mayoría de las críticas que hasta la fecha se han formulado contra el planteamiento narrativo clásico. Sin embargo, como he venido señalando desde el segundo capítulo, hay un tipo de crítica que ha recibido el PNC y que necesita aún ser considerada, pues también puede afectar al NPN. Se trata del posible abuso no justificado de la comparación entre personas y personajes literarios en base a una utilización ambigua del concepto de «narración». En este capítulo espero dar respuesta a este problema y a todas las ramificaciones del mismo que han ido apareciendo a lo largo de las últimas páginas. Para ello en primer lugar mostraré que todas las críticas recibidas se reducen a una supuesta división entre un sentido «trivial» y otro «literario» de narración, de tal modo que los narrativistas sólo estarían justificados para utilizar el término en el primer sentido y no en el segundo, como han estado haciendo. Tras ello argumentaré que esta división entre dos sentidos de narración, aunque está inspirada en un error presente en los trabajos de los narrativistas, no es acertada, y que si se realiza un trabajo de clarificación sobre este aspecto, las críticas que se han esgrimido contra las teorías narrativas de la identidad pierden su objeto. De este modo defenderé que, si se tienen en cuenta las diferencias entre los Yoes, las personas y los personajes literarios, las comparaciones entre estos tipos de entidades son no sólo posibles, sino también productivas.

6.1. Críticas recibidas

Como decía, la principal crítica que han recibido las teorías narrativas con respecto a la comparación que hacen entre personas y personajes literarios, y de la cual derivan todas las demás críticas al respecto, parte de la ambigüedad con la que los narrativistas han utilizado el término «narración», pues a menudo parecen haber supuesto equivocadamente que toda narración es literaria. Podemos denominar a este problema, que fue señalado especialmente por John Christman (2004) y Peter Lamarque (2004, 2007), el *problema de la ambigüedad*.

Tal y como señalan estos autores, aunque existe una relación evidente entre literatura y narrativa –las novelas y los cuentos son ambas cosas–, se trata de dos conceptos distintos, pues ni toda obra literaria es narrativa –la poesía, el teatro y el ensayo no tienen por qué tener ningún elemento narrativo– ni toda narración tiene la complejidad y la intención artística que tienen las obras literarias –un hombre que miente a su mujer sobre lo que ha hecho la noche anterior está creando una narración, pero no una obra de arte–. Al no haber tenido en cuenta este hecho los teóricos del PNC utilizaron el término de forma inapropiada, sacando conclusiones sobre la similitud entre nuestra vida y la vida de los personajes literarios que suponían asumir un concepto de narración mucho más ambicioso que el que sus demostraciones les permitían utilizar (Lamarque, 2004, pp. 402-403). Un ejemplo claro podría ser el de Alasdair MacIntyre, que pasa de proponer que para comprender lo que un hombre está haciendo en su jardín hemos de elaborar una narración que tenga en cuenta sus intenciones y el contexto que le rodea, lo cual parece trivialmente verdadero, a comparar nuestra vida con las obras de Shakespeare (MacIntyre, 1981, pp. 254, 263).

La acusación sin embargo no se reduce a unos pocos autores, sino que la misma confusión ha dominado el debate. Así por ejemplo Jerome Bruner propone una definición extremadamente laxa, y considera que algo es una narración cuando «se ocupa de las vicisitudes de la intención» (Bruner, 1986, p. 29), lo que supone que cualquier acción humana descrita en términos intencionales ya se puede considerar narrativa. Pero esto no impide a Marya Schechtman, que se basa en el pensamiento de Bruner para proponer su propia caracterización, identificar «narración» con «relato», entendiendo por este último término una historia lineal con principio, nudo y desenlace como las que aparecen en las narraciones literarias antes de las innovaciones estilísticas que trajo consigo el siglo XX (Schechtman, 1996, p. 96). Otro ejemplo sería el de Daniel Dennett, quien únicamente demuestra que la conciencia es un producto de la

inmersión del hombre en un contexto lingüístico, lo que considera que es una base suficiente para considerarla también como narrativa y compararla con el método del escritor James Joyce para representar el flujo de conciencia en sus novelas (Dennett, 1991a, pp. 222-227).

El paso entre un sentido mínimo de lo que es una narración al más elaborado que implica una narración literaria ha sido común en prácticamente todos los autores a pesar de que no ofrecen ningún argumento que lo respalde, lo que les ha permitido otorgar un papel a las narraciones que dominan nuestra vida que podría no estar justificado. De hecho, tal y como señala David Velleman, si no se da ese paso lo máximo de lo que pueden dar cuenta dichas narraciones es de acciones tan mínimas como preparar la comida, responder al teléfono o rascarse un picor (Velleman, 2005, pp. 222-223), por lo que en ningún caso serían suficientes para otorgar una unidad a nuestra vida salvo quizá en el sentido trivial de ser todas ellas acciones causadas por un mismo cuerpo físico, para lo cual no necesitamos recurrir al concepto de «narración» (Christman, 2004, p. 702).

De este modo aparece el primer problema derivado de esta crítica y que como vimos fue señalado también por Galen Strawson: el *problema de la trivialidad*. Si tratamos de dar una unidad a lo que los narrativistas parecen estar entendiendo por «narración» cuando quieren demostrar que nuestro pensamiento es narrativo, parece que lo que quieren decir es que algo es una narración cuando está compuesto por frases realizadas en términos intencionales. La vinculación con el lenguaje y con la intencionalidad parecen ser los rasgos definitorios de lo que están entendiendo por narración. Pero si lo máximo que pueden defender es que las personas se construyen narrativamente porque se comprenden a sí mismas mediante frases que contienen términos intencionales, sin duda se trata de una reivindicación trivial que no parece aportar nada. De hecho, al contrario de lo que sugiere Christman, podría ser que esta reivindicación ni siquiera fuese suficiente para aportar una unidad a nuestra vida en el sentido trivial de que todas nuestras acciones están causadas por un mismo cuerpo físico, pues la mayoría de nosotros no articulamos narraciones de ningún tipo para explicar nuestros comportamientos. La mayor parte de nuestra vida nunca es narrada verbalmente de forma explícita, pues no tenemos necesidad de explicar lo que hacemos cuando estamos solos, ni cuando los demás pueden entender por sí mismos nuestras acciones observándonos realizarlas. Es más, a excepción de aquellos que deciden escribir su autobiografía, tampoco intentamos producir una narración más amplia que unifique nuestra vida aunque sea ignorando los aspectos triviales.

Aparece así el segundo problema derivado, y que podemos denominar como el *problema de la explicitud*. Consciente de que las personas rara vez articulan de forma explícita una narración de su vida al completo, Schechtman señalaba en la primera versión de su teoría de la identidad narrativa la necesidad de que las personas tuviesen algún tipo de narración implícita y fuesen capaces de articularla explícitamente si les era solicitada (Schechtman, 1996, pp. 114-119). Pero Lamarque señala que la idea de que las personas tienen una narración implícita de sus vidas esperando a ser narrada, como parece sugerir Schechtman, es incoherente. Tanto en el sentido trivial como en el literario, para que algo sea una narración tiene que haber sido contado de algún modo, pues no existe narración sin texto. En este sentido los personajes literarios dependen ontológicamente de las palabras utilizadas para su descripción, por lo que no existen sin ellas. Dado que las personas existen aunque no articulen de un modo explícito una narración de su vida al completo, asegurar que las personas se construyen narrativamente parece claramente una equivocación (Lamarque, 2004, pp. 404-406).¹

En una línea todavía más radical se ha pronunciado también otro filósofo de la literatura como Gregory Currie, aunque no desarrolla sus argumentos con la profundidad requerida al tratar el tema sólo de forma tangencial. Para Currie, al igual que para Lamarque, las personas no pueden estar constituidas por una narración dado que no son un texto (en su terminología, no son un «artefacto representacional»). Pero en un segundo movimiento más arriesgado asegura que sus vidas tampoco pueden ser el objeto de una narración, la historia narrada (Currie, 2010, pp. 24-25). Aunque algunas personas puedan efectuar narraciones sobre sí mismas, sus vidas serán siempre mucho más complejas que cualquier narración que puedan elaborar, de modo que aunque sus narraciones puedan contener algunos fragmentos de su vida, se caracterizarán sobre todo por los múltiples episodios ignorados y por la gran cantidad de errores cometidos.

Este aspecto nos lleva hasta el tercer problema derivado de la crítica a la comparación entre personas y personajes literarios: el que Fernando Broncano denomina el *problema del escepticismo* (Broncano, 2013, p. 187). Algo curiosamente aceptado por muchos narrativistas es que

1 La importancia de esta crítica quedó reconocida por Schechtman en 2011, cuando intentando responder a Lamarque asume que se trata de un problema que de momento ninguna teoría del Yo narrativo ha sido capaz de contestar satisfactoriamente, por lo que la apunta como una de las áreas a explorar en el futuro (Schechtman, 2011, pp. 406-407, 415).

nuestras narraciones sobre nosotros mismos están en mayor o menor medida ficcionalizadas. Esta postura normalmente es defendida señalando que nuestra memoria autobiográfica no es fiable, pues comúnmente tendemos a reconstruir el pasado del modo que nos sea más beneficioso.² A estas consideraciones a menudo también se une la idea de que el propio carácter narrativo de nuestras historias implica algún tipo de falsificación de los hechos, como si los conceptos de «narración» y «ficción» se implicasen mutuamente (Lamarque, 2004, pp. 397-400). Sin embargo si esto fuese cierto las reivindicaciones de los propios narrativistas en temas morales se verían seriamente debilitadas, pues sin duda hacer depender el éxito ético de algún tipo de falsificación no es lo más deseable (Strawson, 2004, p. 202). Este problema, como vimos en el capítulo dos, era evitado en el PNC ora gracias las narraciones de los otros –que servían de control para que las nuestras no se desviasen de los hechos– ora mediante la postulación de algún tipo de restricción que la realidad debía imponer a nuestras narraciones (Schechtman, 1996, pp. 119-130). La solución en el NPN adoptará formas parecidas, pero será mejor no adelantarnos a nuestros pasos.

Finalmente, el último problema derivado de la comparación entre personas y personajes literarios es el que podemos denominar como el *problema de la interpretación*, y que se refiere al tipo de actitudes interpretativas que es lícito adoptar a la hora de enfrentarse a las narraciones que nos constituyen. Pero para explicar con más detalle en qué consiste el problema es necesario dar un pequeño rodeo

Para demostrar que las vidas de las personas no pueden ser narrativas en el sentido en que lo son las obras literarias sino únicamente en un sentido trivial, Lamarque y Christman ofrecieron de forma independiente dos argumentaciones muy parecidas basándose principalmente en dos tipos de relaciones entre eventos que según ellos sirven para dar unidad a una narración literaria y que sin embargo se encuentran ausentes en nuestras vidas. El primer tipo de relación se basa en el principio teleológico que guía cualquier narración literaria, y que supone que todos sus elementos contribuyen de algún modo a llegar a un punto culminante

2 Sobre la evidencia del carácter reconstructivo y revisionista de la memoria véase lo dicho a propósito de las críticas al concepto de cuasi-memoria en el capítulo uno. A lo señalado entonces también pueden añadirse opiniones como la del psicólogo Michael Gazzaniga, para quien la forma en que nuestro cerebro recuerda nuestra vida implica que «toda autobiografía es irremediamente fabulatoria» (Gazzaniga, 1998, p. 20). Por otro lado, Strawson señala acertadamente que los procesos revisionistas no tienen por qué darse únicamente en nuestro propio beneficio (*cf.*: Strawson, 2004, pp. 201-204).

en la historia. La segunda relación es la unidad temática subyacente en una obra literaria y que le da su valor artístico.³

El principio teleológico nos permite, cuando se aplica a una obra literaria, otorgar sentido a todos los elementos de la narración en función del papel que desempeñan en la consecución del objetivo de la misma. Así por ejemplo podemos entender la muerte de un personaje como una necesidad para que la obra tenga un carácter trágico. Sin embargo, si intentamos aplicar dicho principio a nuestras vidas incurrimos en una falsificación, pues no es cierto que nuestra vida tenga un sentido predeterminado al que todas nuestras experiencias tengan que contribuir (Lamarque, 2007, p. 131). No podemos explicar nuestra decisión de aprender un idioma como una necesidad estructural del argumento para justificar una futura relación amorosa con una extranjera que será el amor de nuestra vida. Nuestra vida requiere de explicaciones causales, no teleológicas. Aprendemos un idioma porque queremos realizar un viaje, no para poder cumplir nuestro destino. Por otro lado, y al contrario de lo que sucede con las narraciones literarias, la mayoría de las personas carecen de un único objetivo al que todas sus experiencias contribuyan de modo que sea posible dar una explicación teleológica *a posteriori*, sino que persiguen a la vez distintas cosas de forma independiente. Así podemos tener por ejemplo experiencias dirigidas a la consecución de un trabajo, a tener una revelación espiritual o a disfrutar de una relación sentimental sin que dichos aspectos de nuestra vida tengan ninguna relación entre sí. Pero si aceptamos la existencia de múltiples objetivos, entonces una explicación teleológica no será capaz de unificar la historia de nuestra vida como sí hace con las narraciones literarias (Christman, 2004, p. 704).

El otro aspecto que permite dar unidad a las narraciones literarias según Christman y Lamarque es el tema que abordan. El tema puede ser una idea o un sentimiento que se quiere transmitir o analizar a lo largo de toda la narración –el amor, el paso del tiempo, la complejidad de las relaciones entre una madre y sus hijas, etc.–, y que permite unir episodios aunque estén totalmente desconectados desde un punto de vista causal. Por ejemplo una novela puede constar de dos historias paralelas unidas por reflejar una misma problemática o por elementos simbólicos, como sucede en *Las palmeras salvajes* de William Faulkner. El valor artístico de una

3 Lamarque y Christman prefieren tomar como modelo de narración literaria obras de corte tradicional, igual que hacen los narrativistas, evitando así las complicaciones de caracterizar las obras de la literatura moderna y postmoderna que en ocasiones no presentan una unidad en ninguno de los sentidos ofrecidos por estos autores. Se trata de un procedimiento discutible, pero volveré sobre ello más adelante.

obra literaria depende en gran medida de la profundidad y la perspicacia con que aborde su tema, por lo que se trata de un elemento de gran importancia en este tipo de narraciones. Sin embargo resulta evidente que las vidas de las personas no tienen un tema en este sentido, por lo que no podemos comparar las narrativas autobiográficas con las literarias. Aunque algunos biógrafos pueden intentar darle unidad al material con el que cuentan buscando un tema común a todos o a la mayoría de los episodios de la vida de alguien, al hacerlo se están alejando de la realidad ficcionalizándola, y si persisten demasiado en ello finalmente acabarán escribiendo una novela y no un relato ajustado de la vida de una persona real (Lamarque, 2007, p. 132).

No obstante podemos darnos cuenta de que las unidades teleológica y temática de una obra literaria dependen en gran medida de la labor hermenéutica del intérprete de la misma, por lo que siempre es posible que distintos intérpretes encuentren temas diferentes en una misma obra o que un crítico imaginativo encuentre una explicación teleológica en una narración extremadamente inconexa (*cf.*: Christman, 2004, pp. 705-706). En este sentido podemos recordar la respuesta que Schechtman dio a las críticas de Lamarque –por algún motivo ignora las de Christman pese a su similitud– inspirándose en Paul Ricoeur, y señalando que el problema reside en que Lamarque adopta la visión de un crítico cuando se enfrenta a una obra literaria, pero cuando lo hace con la narración de una vida adopta la visión de un personaje. Sin embargo las personas tienen que adoptar al mismo tiempo los roles de crítico, autor y personaje con respecto a sus propias vidas, por lo que pueden tener distintos tipos de interpretaciones de las mismas (Schechtman, 2011, pp. 411-414). En cualquier caso parece que ni siquiera adoptando el rol de un crítico con respecto a nuestras vidas estamos justificados para darles una unidad teleológica o temática del modo que sí se lo podemos dar a una obra literaria, por lo que la respuesta de Schechtman revela finalmente el sentido de lo que hemos denominado como el problema de la interpretación, y que como decíamos se refiere a qué tipo de actitudes interpretativas es lícito adoptar con respecto a las narraciones de nuestras vidas.

6.2. Solucionando el problema

Las críticas que hemos visto en el apartado anterior sin duda tienen una base fundamentada, pues el uso de los narrativistas del concepto de «narración» es decididamente ambiguo. Esta ambigüedad provoca que las teorías narrativas sean vulnerables además a los cuatro problemas

derivados que hemos señalado: el de la trivialidad, el de la explicitud, el del escepticismo y el de la interpretación. Sin embargo creo que las críticas de Christman, Lamarque, Strawson o Currie no aciertan a dar en el blanco. Aunque identifican correctamente el problema principal, que es el hecho de que la ambigüedad con la que se utiliza el término «narración» provoca que se atribuyan a las narraciones «triviales» propiedades que sólo tienen las «literarias», no son capaces de detectar el sentido de «narración» que es relevante para las teorías de la identidad narrativa. Si lo hubiesen hecho se habrían dado cuenta de que los problemas de la trivialidad y de la explicitud no se seguirían del problema de la ambigüedad, mientras que los del escepticismo y de la interpretación podrían ser resueltos sin mayores complicaciones. En consecuencia para poder efectuar comparaciones apropiadas entre los Yoes, las personas y los personajes literarios los narrativistas sólo tienen que desarrollar el sentido de «narración» relevante para sus teorías y tener en cuenta las diferencias que existen entre las narraciones que dan lugar a nuestra existencia y las que dan lugar a la existencia de los personajes literarios. Espero que las páginas que siguen sirvan para aclarar este problema y abrir al tiempo una nueva línea de investigación en cuanto a la comparación entre personas y personajes se refiere.

Pero antes de responder a las críticas, un punto que merece la pena destacar es que éstas fueron realizadas contra el PNC, por lo que no tienen cuenta la diferenciación que se hace en el NPN entre Yoes y personas. Este aspecto no tiene realmente importancia en lo que se refiere a la resolución del problema de la ambigüedad, pues tanto los Yoes como las personas se entiendan a sí mismas (y a otros Yoes y personas) de forma narrativa en el mismo sentido. Tampoco lo tendrá en consecuencia para los problemas de la trivialidad y de la explicitud –que dejan de tener efecto en las teorías narrativas una vez solucionado el problema de la ambigüedad– ni para el de la interpretación. Sin embargo el problema del escepticismo deberá ser abordado de forma distinta cuando se hable de una u otra entidad.

6.2.1. Hacia un concepto apropiado de narración

Como decía, los críticos de las teorías narrativas no interpretan adecuadamente la ambigüedad presente en las mismas, lo que les lleva a equivocar sus argumentaciones. Y el motivo es que no se dan cuenta de que los sentidos «trivial» y «literario» de narración que se pueden definir a partir del uso que hacen del término los narrativistas no son entendibles como versiones más o menos elaboradas de un mismo fenómeno. El

sentido «literario» de narración no es la suma del sentido «trivial» más una serie de requisitos extra, como parecen sostener. Aunque es cierto que no toda narración es literaria y que una narración literaria puede comprenderse como la suma de una narración simple y de una serie de requisitos añadidos, en el caso que nos ocupa esto no sucede. Las definiciones de narración en un sentido «trivial» o un sentido «literario» que manejan o parecen manejar los narrativistas están elaboradas desde marcos teóricos distintos, aspecto que Christman y Lamarque ignoran.⁴

El problema de los narrativistas es que demuestran que las personas son narrativas según un marco teórico –«cognitivo»– de lo que debe entenderse por «narración» y de ahí dan un salto injustificado al suponer que también deben ser narrativas en el sentido en que los personajes literarios lo son según otro marco teórico –«formalista»–. Los anti-narrativistas por su lado parten de ese marco formalista para definir lo que significa «narración» en el sentido «literario» y dentro de ese mismo marco eliminan las características que convierten en «literaria» a una narración para definir lo que significa el término en el sentido «trivial». El problema entonces de los anti-narrativistas es suponer que su definición formalista de «narración» en el sentido «trivial» es equivalente a la definición cognitiva que los narrativistas hacen del término. Pero al hacerlo ignoran que un marco no es reducible al otro. Para mostrar esto podemos fijarnos en las relaciones que se dan entre los conceptos de «narración» que manejan Bruner, Schechtman y Lamarque respectivamente.

Como ya hemos dicho, para Bruner nuestro pensamiento es narrativo si se realiza en términos intencionales (Bruner, 1986, p. 29). Pero cuando dice esto no quiere significar que una frase sea narrativa cuando contiene verbos como «intentar» o «desear». Lo que está intentando transmitir es que existe un modo primitivo de pensamiento que se puede calificar como «narrativo» y que es el que utilizamos para comprender el comportamiento de las personas.⁵ En un momento llega incluso a preguntarse si el

4 En honor a la verdad, Christman, inspirándose en Ricoeur, ve el problema de un modo parecido al mío. Sin embargo llega a la conclusión de que el marco desde el que los narrativistas definen el sentido «trivial» de «narración» no debería ser denominado «narrativo», pues puede llevar a confusión (Christman, 2004, p. 709).

5 Es importante señalar la vinculación de este modo narrativo del pensamiento con la temporalidad (ya mencioné anteriormente la influencia de Ricoeur en el pensamiento de Bruner). A este respecto podemos decir que el modo narrativo del pensamiento es también el que utilizamos para humanizar el tiempo y darle sentido. Es el que está involucrado en la conversión del monótono sonido del reloj en la proto-narración «tic-tac» (Kermode, 1966, pp. 51-54). Por otro lado, el modo en que se relacionan entre sí la experiencia de la temporalidad, la

mecanismo que utilizamos para reconocer que alguien es una persona no será precisamente el comprobar si se puede convertir en una narración de alguna clase (Bruner, 1997, p. 152). Para Bruner por tanto algo es narrativo no cuando da lugar a frases de uno u otro tipo, sino cuando es interpretado utilizando el modo narrativo del pensamiento.⁶

Por su parte Schechtman, aunque asegura inspirarse en el concepto de «narración» de Bruner, dice que las personas construyen su identidad mediante el relato de su vida, entendiendo «relato» como una construcción lingüística con unos rasgos formales determinados que debe estar codificada explícitamente en un texto (motivo por el que se ya desde el principio de su teoría se preocupa tanto por el problema de la explicitud) (Schechtman, 1996, p. 96). De este modo podemos ver cómo se produce en las teorías narrativas la confusión de la que hablábamos, pues Schechtman se sirve de la definición cognitiva de Bruner para dar injustificadamente el salto a una definición formalista.

Finalmente Lamarque parte del supuesto de que el prototipo de narración son también las narraciones literarias de corte clásico en las que está pensando Schechtman. Para él narrar significa contar una historia, por lo que no puede haber una narración sin un texto que la contenga, ya sea de forma oral o escrita. Aunque el sustantivo «narración» puede referirse a la historia narrada, al acto de narrar o al propio texto, los tres significados tienen en común la idea de que no puede haber una narración sin un acto explícito que implique narrar (Lamarque, 2004, pp. 394-395). En este sentido se debe tener en cuenta que incluso cuando «narración» se refiere a la historia narrada, el término implica necesariamente la existencia de un acto en el que se haya contado esa historia. Aunque en este significado del término puede haber distintas versiones de una misma narración, así como por ejemplo hay muchas versiones del mito de Fausto, no hay una historia (narración) de Fausto que sea independiente o previa a la primera versión explícitamente contada del mito. Siendo así, nuestra vida no puede ser narrativa si no ha sido contada explícitamente.

Armado con esta definición formalista de lo que son las narraciones, Lamarque procede a enumerar los rasgos mínimos que debe tener una narración para ser tal: (1) debe ser el producto de una acción humana, no

comprensión de la intencionalidad humana y la narratividad ha sido abordado recientemente desde las ciencias cognitivas por Erica Cosentino (2011).

6 En este sentido podemos señalar cómo desde el interés por los estudios literarios la poética cognitiva recurre a las ideas de Bruner para defender que «tanto la narración como la metáfora son antes que nada modalidades del pensamiento o, en otros términos, estrategias mentales orientadas hacia el conocimiento de la realidad» (Garrido Domínguez, 2011, p. 166).

algo que se dé en la naturaleza; (2) debe contener dos sucesos que estén conectados de alguna forma; (3) debe introducir una dimensión temporal en la relación entre los sucesos; (4) debe poder ser identificada por sus rasgos formales ya desde el nivel de las frases individuales. En este sentido, según Lamarque, la frase «el sol brilló y la hierba creció» es narrativa, pues cumple todos estos requisitos. Sin embargo «el niño golpeó la pelota y la pelota fue golpeada por el niño» no lo es, pues no presenta dos eventos que se relacionen entre sí, sino un único evento formulado de dos formas distintas. Esta definición mínima de lo que es una narración se convierte así en el modelo según el cual se debe entender el sentido «trivial» que según Lamarque utilizan los narrativistas, lo que provoca que el sentido cognitivo de narración ofrecido por Bruner sea interpretado (como vimos en el apartado anterior) como una variación del sentido formalista: algo es una narración cuando está compuesto por frases que contengan términos intencionales. La perversión del sentido original que Bruner pretendía darle al término es evidente.⁷

Suponer que el concepto formalista de narración «literaria» es reducible a la suma de un concepto formalista «trivial» más una serie de requisitos no es un error. Los errores son suponer que el concepto formalista «trivial» es equivalente al concepto cognitivo ofrecido por Bruner (como hacen los anti-narrativistas) y suponer que el concepto cognitivo puede utilizarse para llegar al concepto formalista «literario» (como hacen los narrativistas). Si se sostiene una postura cognitiva –como sin duda los narrativistas deberían hacer⁸– y se piensa que algo es narrativo cuando es interpretado de cierto modo, entonces no se puede acudir a un marco formalista para definir las narraciones literarias. Es

7 Es interesante leer desde esta perspectiva las críticas que Bernard Williams hizo a la postura de MacIntyre, pues implican el mismo tipo de interpretación formalista del sentido cognitivo de narración. Para Williams lo que MacIntyre debe querer decir cuando señala que la estructura narrativa de las acciones humanas es anterior a las narraciones que se pueden ejemplificar a partir del trabajo de los novelistas es sin duda que una narración no-artística es previa a una narración artística. La otra posible interpretación –que la estructura narrativa es previa a cualquier narración explícita– le parece sencillamente un sinsentido (Williams, 2007, p. 306).

8 Aunque lamentablemente no todos los narrativistas aceptan que el sentido relevante de «narración» deba ser el cognitivo. Por poner un ejemplo reciente, si se analiza la postura de Mary Jean Walker –quien hace apenas unos meses intentó dar una respuesta al problema del escepticismo– podemos ver que continúa pensando que las personas se comprenden a sí mismas narrativamente en un sentido formalista (Walker, 2012, pp. 64-65).

necesario recurrir a un marco como el de la estética de la recepción o el de la poética cognitiva.⁹

La estética de la recepción es una corriente en teoría de la literatura que, desde finales de los años sesenta del siglo XX se desarrolló sobre todo en Alemania (*vid.* Mayoral, 1987; Warning, 1975). Su principal reivindicación fue que para comprender el fenómeno literario es de vital importancia atender al papel del lector, aspecto ignorado hasta aquel momento por la teoría y la crítica literarias. Con esta idea en mente, desde la estética de la recepción se defiende que la obra literaria no se realiza en el texto elaborado por el autor, sino en el proceso de interpretación realizado por el lector.¹⁰ El texto no es más que un conjunto de símbolos impresos en un libro, y no es hasta que el lector les da un significado que éstos se convierten en una obra literaria.¹¹ Sin embargo, y esta es otra de las principales reivindicaciones de la estética de la recepción, los textos nunca tienen un significado unívoco. Aunque el autor impone límites a lo que su texto puede significar, es una condición incluso del relato más sencillo omitir partes de la historia (Iser, 1972, p. 222). Nada puede contarse en su totalidad pues siempre existe un contexto más amplio en el que podría encuadrarse un suceso para obtener una mayor comprensión. De este modo es una característica esencial de toda narración el poseer partes no narradas. Todo texto posee espacios en blanco, lugares de indeterminación en los que el texto guarda silencio (Ingarden, 1968, pp. 36-37). Pero como el lector necesita otorgar coherencia a lo que dice el

9 La terminología no debe inducirnos a confusión y hacernos pensar que la poética cognitiva debería ser más relevante que la estética de la recepción en la búsqueda de un concepto de narración más adecuado para las teorías narrativas de la identidad. Como señala Antonio Garrido Domínguez, la orientación cognitiva está muy acentuada tanto en una como en otra corriente (Garrido Domínguez, 2011, p. 160). En este sentido en las próximas páginas me basaré sobre todo en la estética de la recepción, aunque podría ser igual de provechoso hacerlo en la poética cognitiva o en ambas al tiempo. Por otro lado no podemos dejar de señalar aquí que también Ricoeur se inspiró ampliamente para escribir *Tiempo y narración* en autores de la estética de la recepción, especialmente Wolfgang Iser (Garrido Domínguez, 2004, p. 107).

10 Para ser más exactos habría que decir que la obra no reside ni en el texto elaborado por el autor ni en la concretización individual del lector, sino en el punto de convergencia entre ambos. Un punto que nunca puede localizarse con precisión, pues cada lector individual lo sitúa en un lugar diferente, como veremos más adelante (Iser, 1972, pp. 215-216).

11 En este sentido se puede interpretar la postura de Peter Kivy, para quien toda la literatura debería ser considerada como una forma de arte dramático que el lector representa al poner voz al texto durante la lectura (*vid.* Kivy, 2010).

texto para poder comprenderlo, al leer rellena dichos huecos en blanco mediante su imaginación realizando un proceso de concretización que es distinto para cada individuo (Iser, 1972, pp. 228-230). Las teorías literarias anteriores, como el formalismo, habían centrado su atención sobre todo en obras literarias de corte clásico en las que los lugares de indeterminación son menos numerosos, por lo que no les prestaron atención. Pero como desvela la estética de la recepción, desde el siglo XVIII han ido incrementándose los lugares de indeterminación en las obras literarias (Iser, 1972, p. 223). Llegando en el siglo XX hasta el extremo en que se hace evidente que el carácter no sólo literario, sino también ficticio e incluso narrativo de una obra no está determinado por el texto, sino por la concretización que el lector hace del mismo (*cf.* Carr, 1998, p. 342). Como ejemplo podemos pensar en el relato «El índice» de J. G. Ballard, incluido en su libro *Fiebre de guerra*. Dicho relato consiste únicamente en el índice temático y onomástico de la supuesta autobiografía perdida de Henry Rhodes Hamilton (HRH), por lo que no encontramos en él más que una colección de nombres y sucesos ordenados alfabéticamente:

- Acapulco, 143
- Acton, Harold, 142-7, 213
- Alcázar, Asedio del, 221-5
- Alimenticia, HRH paga la pensión, 172, 247, 367, 453
- Anaxágoras, 35, 67, 69-78, 481
- Apollinaire, 98
- Arden, Elizabeth, 189, 194, 376-84
- Autobiografía de Alice B. Tocklas, La (Stein), 112
- Aviñón, lugar de nacimiento de HRH, 9-13; vacaciones durante la infancia, 27; investigación en el Instituto Pasteur de Oftalmología, 101; tentativas de restablecer el antipapado, 420-35 (Ballard, 1990, p. 210)

Desde una postura formalista, el relato de Ballard es difícilmente clasificable como un texto narrativo, pues no tiene los rasgos necesarios. Se señala a menudo que una mera recolección de hechos, como la que se puede encontrar en unos anales, no constituye una narración, pues es necesario que se establezcan vínculos causales o explicativos entre esos hechos (*vid.* Vice, 2003, p. 95; White, 1980). Pero en el relato de Ballard esto no sucede. Si nos fijamos en la cita que hemos reproducido, la relación por ejemplo entre Acapulco y Harold Acton se insinúa en cierto sentido, pues ambas entradas refieren a las mismas páginas de la autobiografía. Sin embargo no se dice de forma explícita que tipo de relación guardan entre sí –¿HRH y Harold Acton hablaron de Acapulco? ¿Se fueron de viaje a Acapulco juntos? ¿Se cruzaron accidentalmente en la

playa?-. Y si vamos todavía más lejos, la relación entre Anaxágoras y Elizabeth Arden es completamente inexistente.

Si adoptamos por el contrario la postura de la estética de la recepción podemos ver que entre dos entradas cualesquiera del índice existe al menos una conexión: la asunción que hace el lector de que todas ellas deben tener algún tipo de relación con la vida de HRH. Los lugares de indeterminación del relato son tan numerosos que el texto de Ballard parece no ser más que una sucesión de datos inconexos. Sin embargo un hipotético lector que se enfrentase al texto con esas expectativas no tendría ningún interés en continuar su lectura. Como hemos dicho, el lector necesita otorgar sentido y coherencia a los textos que lee para comprenderlos, y para ello tiene que asumir la tarea de concretizar los lugares de indeterminación. Lo que el ejemplo de Ballard pone de relieve es que este acto de concretización supone interpretar el texto utilizando el modo narrativo del pensamiento, pues sólo así podrá darle una unidad. Lo que convierte entonces al relato de Ballard en una narración no son sus rasgos formales, sino el modo de pensamiento con el que el lector lo interpreta. Al suponer que se trata de un relato, el lector es capaz de establecer las relaciones necesarias entre las palabras del texto para dotarlas de una unidad narrativa. Desde esta perspectiva los narrativistas no deben preocuparse entonces por buscar las condiciones formales mínimas que debe tener el relato de una vida para ser narrativo, sino de los mecanismos psicológicos y sociales implicados en el proceso de dar un sentido narrativo a cualquier texto, acción o serie de acciones (Christman, 2004, pp. 709-711).

De este modo podemos ver que, al solucionar el problema de la ambigüedad y utilizar el concepto apropiado (cognitivo) de «narración», los problemas de la trivialidad y de la explicitud también desaparecen. En cuanto al primero, tal vez sea trivial considerar que cuando hablamos sobre nosotros lo hacemos con frases que contienen términos intencionales. Pero defender que existen dos modalidades primitivas del pensamiento como hace Bruner, y que una de éstas la utilizamos específicamente para entendernos a nosotros mismos no es para nada trivial. De hecho, como señala Anthony Rudd, existen modelos científicos alternativos para explicar las acciones humanas ante los que es importante defender la postura narrativa (Rudd, 2007, p. 63). Sin embargo la acusación de trivialidad podría extenderse al concepto cognitivo de «narración», como parece hacer Strawson. La idea de Strawson es que las narraciones mediante las que comprendemos nuestras acciones son discretas, independientes unas de otras, de tal modo que su existencia no sería relevante para dar unidad a la totalidad de nuestra vida. Su ejemplo

sobre la narrativa que implica preparar café sería una muestra de que para él existen infinidad de narraciones discretas de este tipo en cada día. Sin embargo, como también señala Rudd –y de forma análoga a lo que hemos dicho sobre las narraciones literarias desde la estética de la recepción–, al igual que los sucesos de nuestra vida sólo son comprensibles en el marco de pequeñas narraciones como la necesaria para preparar un café, a su vez estas pequeñas narraciones sólo son comprensibles en el marco de narraciones más extensas (Rudd, 2007, p. 64). Una narración concreta sobre preparar café se encaja dentro de un contexto más amplio –¿es el típico café que me tomo por las mañanas? ¿es un café que preparo para una visita inesperada? ¿lo estoy preparando para escaparme de la rutina del trabajo?–, y ese contexto a su vez sólo puede ser comprendido en forma narrativa. En este sentido, la inteligibilidad de un suceso no es una cuestión de todo o nada, sino una cuestión de grado, pues siempre es posible enmarcarlo en contextos cada vez más amplios para aumentar nuestra comprensión del mismo. Los sucesos de nuestra vida no se encuentran aislados entre sí, y del mismo modo las narraciones mediante las que los comprendemos tampoco lo están. Si en ocasiones lo parece sólo es porque por motivos pragmáticos decidimos no seguir ampliando indefinidamente el contexto, pues la comprensión que tenemos de los sucesos ya es suficiente para nuestros propósitos. Siendo así el modo narrativo del pensamiento es capaz de darle una unidad a toda nuestra vida en un sentido profundo que va mucho más allá de la constatación del hecho de que todas nuestras acciones fueron realizadas por nuestro cuerpo.

En cuanto al problema de la explicitud, lo dicho hasta el momento revela como evidente que no tiene objeto cuando se está utilizando el sentido «cognitivo» de narración. Lo importante para decir que nuestra vida se constituye de forma narrativa no es el hecho de si efectivamente hemos elaborado una narración al completo de nuestra existencia, sino que para comprenderla hemos de utilizar el modo narrativo del pensamiento (*cf.* Bruner, 1997, p. 155). Que las narraciones que produzcamos de forma explícitamente textual nunca aborden más que una pequeña parte de nuestra vida no tiene importancia, pues a diferencia de lo que autores como Lamarque o Currie parecen creer, éste es precisamente uno de los rasgos característicos de toda narración. Las narraciones literarias tampoco pueden incluir la vida al completo de sus personajes, ni siquiera cuando se presentan bajo la forma de una biografía. Seguramente la novela de Laurence Sterne *Vida y opiniones del caballero Tristram Shandy* sea el ejemplo más apropiado a este respecto, pues su tema principal parece ser precisamente la inenarrabilidad de la vida humana. En esta novela del siglo XVIII el narrador revela su intención de contarnos con todo lujo de

detalles su vida hasta alcanzar el momento de la escritura de su obra. Sin embargo desde el principio su proyecto se revela como imposible: para poder contar cómo nació se ve en la necesidad de contar cuáles eran las costumbres sexuales de sus padres, el contenido del contrato prematrimonial que firmaron y que especificaba las condiciones que debía reunir su lugar de nacimiento, que sólo existía una comadrona en la región y que el motivo de que fuese así no era otro que la generosidad del párroco de la zona, las discusiones entre los intelectuales de la época sobre si era posible bautizar a los niños aún no nacidos... El resultado, en definitiva, es que tras doscientas páginas el narrador aún no ha sido capaz de contar su natalicio. Al igual que el personaje de Sterne, nosotros tampoco podremos contar nunca nuestra historia. Pero aunque no lo hagamos, para comprender nuestra vida deberemos hacerlo utilizando el mismo modo narrativo del pensamiento que utilizamos para comprender una novela.

6.2.2. Escepticismo y negociaciones narrativas

Como señalé al principio de este apartado, las críticas que hemos estado viendo en este capítulo fueron realizadas antes de que el NPN distinguiese entre Yoes y personas. Este hecho no era relevante para responder a los problemas de la ambigüedad, de la trivialidad y de la explicitud, pues considero que tanto los Yoes como las personas se comprenden a sí mismas en los términos narrativos que hemos expuesto en el anterior epígrafe. Sin embargo sí tiene importancia para abordar el problema del escepticismo, que será el objeto del presente.

En cuanto a los Yoes, el problema del escepticismo es sin duda una amenaza. Como vimos en el capítulo anterior, existe un acceso privilegiado al Yo desde la primera persona, de modo que si alguien asegura que ya no es el mismo Yo que habitaba su cuerpo tiempo atrás, no hay argumento que pueda rivalizar con la percepción subjetiva de dicho Yo. De este modo, además de a los problemas derivados de interpretaciones inválidas sobre lo que alguien es en tanto Yo que ya señalé en aquel momento en una nota al pie, esta situación evidentemente abre la puerta a la posibilidad de que alguien mienta consciente o inconscientemente sobre su propio Yo. Podría decirse que quien miente no está situándose al margen de quien es, sino determinando su existencia de un modo específico: como mentiroso (Cruz, 2005, p. 23). Sin embargo esto no resuelve el problema del escepticismo, cuya cara más reconocible es la duda sobre cómo saber si alguien está mintiendo sobre su Yo.

Seguramente el problema del escepticismo no tenga una solución adecuada en lo que a los Yoes se refiere. Aparte de la consigna de que para constituirse en tanto Yo es necesario hacerlo a través de un mundo válido como señalábamos en el capítulo anterior –consigna que tiene mucho que ver con la restricción que la realidad debe imponer a nuestras narraciones de la que hablaba Schechtman (1996, pp. 119-130)–, parece que no podemos establecer ningún criterio para determinar si un Yo está mintiendo sobre su propio ser o no. Lo que sí podemos defender sin embargo es que este problema no tiene ninguna relación con el hecho de que los Yoes se construyan narrativamente. Como bien señala Lamarque en sus críticas, los conceptos de «narración» y de «ficción» no se implican uno al otro. Podemos tener narraciones perfectamente válidas sobre sucesos reales. De hecho, aunque sólo podamos conocer los hechos relativos a los Yoes en forma de narraciones y que por tanto en cierto sentido las narraciones fabriquen los hechos de los que hablan,¹² esto no significa que no haya mejores y peores narraciones de una misma situación, al igual que hay mejores y peores interpretaciones de la REALIDAD. Lo que esto implica es que el único modo de saber si una narración es acertada o no es compararla con otras narraciones y no con los supuestos hechos objetivos (Rudd, 2007, p. 66). En este sentido aunque el problema del escepticismo siga estando vigente para los Yoes, debemos aceptar que éste no tiene ninguna relación con el hecho de que se construyan de forma narrativa.¹³

12 Esta es la línea de respuesta adoptada por David Carr, para quien tampoco tiene sentido la oposición entre nuestras narraciones y una REALIDAD objetiva previa a las mismas. La realidad en la que vivimos es una realidad humanizada y construida narrativamente, por lo que el recurso a un supuesto sustrato previo es un error (Carr, 1986).

13 En su análisis del problema del escepticismo, Walker señala que existen tres estrategias de respuesta. La primera es la que adoptan Schechtman y MacIntyre, que tiene que ver con las restricciones que imponen la realidad y los otros a nuestras narraciones. La segunda consiste en señalar que aunque nuestra memoria autobiográfica no es fiel en lo que se refiere a recuerdos concretos, sí suele serlo cuando se refiere al sentido global de los sucesos, por lo que se puede decir que existen distintos niveles de verdad. La tercera, que se puede considerar un desarrollo de la anterior, sería la que adoptaría por ejemplo Rudd, señalando que las verdades que tienen que ver con nuestras vidas sólo se pueden capturar en forma narrativa, por lo que el problema del escepticismo en última instancia no tendría relación con el hecho de que nos comprendamos a nosotros mismos de forma narrativa (Walker, 2012, pp. 70-73). Se puede comprobar cómo mi propuesta de solución enlaza estas tres estrategias de respuesta.

Por su parte en el NPN las personas pueden evitar el problema del escepticismo a través de lo que en el capítulo cinco llamé «negociaciones narrativas». La idea básicamente es la misma que les permitía a MacIntyre y Ricoeur evitar el problema en el PNC. Las narraciones que otros efectúan sobre nuestra vida sirven de control para evitar que nuestra narración se convierta en una novela de ficción. Sin embargo el concepto está sujeto al menos a un problema que no abordé en el capítulo anterior, pues para aclararlo de un modo apropiado era necesario saber primero qué se entendía por «narración» en este contexto. Una vez solucionado este tema, me gustaría profundizar un poco más en lo que va implicado en la idea de «negociación narrativa» con la ayuda de las herramientas que nos ofrece la teoría de la literatura.

La principal objeción que se le puede hacer a la idea de que nuestra identidad se construye mediante una negociación narrativa parte del hecho de que las narraciones que existen sobre una misma persona a nivel social son múltiples y con frecuencia contradictorias entre sí. En este sentido si consideramos el conjunto de las historias que nos conforman como si fuese una sola narración, ésta presenta múltiples lugares de indeterminación que deben ser concretizados mediante negociación. Parece surgir así una aporía. Es necesario concretizar la narración, pues ¿cómo podríamos si no defender que nuestra identidad es determinada? Pero si la concretizamos, ¿cómo sabemos que lo estamos haciendo del modo apropiado?

En primer lugar, para entender la naturaleza de esta narración múltiple que nos constituye puede ser de utilidad recurrir a los conceptos de «polifonía» y «dialogismo» elaborados por Mijail Bajtin. Para este teórico de la literatura ruso las novelas tienen un carácter eminentemente social, de forma que en su seno se encuentran toda una serie de voces que presentan sus propias formas alternativas de ver el mundo. No se trata de que los personajes aparezcan como personas vistas a través de la conciencia única del autor, como diría la crítica tradicional, sino que para Bajtin estas voces se presentan como el fruto de conciencias bien diferenciadas con su propia voz personal (Bajtin, 1961, p. 327). De este modo Bajtin entiende la novela como un texto polifónico. Pero las voces de esta polifonía no se encuentran yuxtapuestas, sino que apelan unas a las otras. Dialogan entre sí para mayor esclarecimiento mutuo al tiempo que compiten para alcanzar la conciencia del lector (Bajtin, 1975, p. 164). Los paralelismos con la narración que nos constituye en tanto personas son más que notables (*vid.* Lodge, 2003, pp. 100-101).¹⁴

14 No podemos dejar de señalar que ya desde la psicología y la sociología se había recurrido a las ideas de Bajtin para explicar algunos aspectos del Yo.

En consecuencia este tipo de textos polifónicos se prestan a interpretaciones diversas, pues las distintas voces provocan lugares de indeterminación. Existen distintas visiones de un mismo problema y es el lector el que debe tomar partido y concretizar el texto del modo que considere apropiado.¹⁵ Sin embargo, como señalaron McKoon y Ratcliff, el lector no necesita concretizar absolutamente todos los lugares de indeterminación del texto, sino sólo aquellos que o bien son fácilmente decidibles o bien son necesarios para su propia interpretación (*vid.* McKoon y Ratcliff, 1992).

Si entendemos la narración que nos constituye de este modo, la aporía empieza a disolverse. En primer lugar podemos ver que la psicología social de roles tiene mucho que decir sobre la identidad personal, pues en la mayoría de las ocasiones las personas se comportan de distinto modo en cada situación dando lugar a narraciones divergentes sobre su carácter (*vid.* Gergen, 1991). Sin embargo normalmente no nos vemos forzados a concretizar de modo absoluto la identidad de una persona. Pueden existir narraciones opuestas sobre un mismo individuo siempre y cuando dichas narraciones no entren en conflicto debido a una situación puntual. El ejemplo más evidente es el de nuestra supervivencia que ya comentábamos en el capítulo anterior, aunque existen otros como los momentos en los que distintas esferas de nuestra vida entran en contacto. Sólo en esos casos la negociación narrativa que nos constituye se ve en la necesidad de interpretar las distintas voces de nuestro relato para concretizarlo. De este modo podemos ver cómo los lugares de indeterminación de la narración que nos constituye no impiden que nuestra identidad sea determinada, pues la narración se concretiza cuando es necesario.

El problema entonces es cómo podemos saber si la negociación narrativa que concretiza una indeterminación lo hace del modo apropiado.

Especialmente porque, aunque no me haya detenido en ello, los Yoes también se constituyen mediante un tipo de negociación narrativa interna que establece qué queda fuera o dentro de esos Yoes fluidos y amorfos de los que hablaba en el capítulo anterior (*vid.* Gagnon, 1992, pp. 229-233; Rodríguez González, 2003, pp. 150-153; Shotter, 1993, pp. 8-14).

- 15 En este sentido podemos señalar que la identificación del lector con un personaje adquiere una importancia decisiva, pues según el lector se incline hacia uno u otro personaje la interpretación de la novela será distinta. Un ejemplo de esta situación se puede encontrar en los análisis que yo mismo he realizado sobre las historias de Peter Pan, en las que la interpretación cambia radicalmente según con qué personaje se identifique el lector. Así, mientras que si nos identificamos con Wendy la lectura que haremos será la propia de una novela de aventuras, si lo hacemos con el capitán Garfio nos encontraremos con una historia de terror (Muñoz Corcuera, 2012a, 2012b).

Pues si pensamos en un juicio penal, parece evidente que siempre es posible que la negociación narrativa que tiene que decidir sobre el futuro de una persona se equivoque y acabe enviando a la cárcel a un inocente. Ante esta objeción es necesario aclarar que si todavía queda alguien que no acepte el resultado de una negociación –que en el ejemplo que nos ocupa puede ser el propio acusado–, entonces la negociación narrativa no ha terminado, aunque se incline en un determinado sentido. De hecho las negociaciones narrativas a menudo quedan sin concluir, de modo que la indeterminación que les dio lugar no queda concretizada de un modo absoluto, sino sólo reducida. Sin embargo el hecho de que una negociación se incline en un determinado sentido –como sucede por ejemplo en un juicio en el que el acusado es condenado pero no acepta su culpabilidad– crea realidad futura: desde ese momento en adelante la persona en cuestión será tratada como «culpable», esto es, como responsable del suceso del que se le acusó, aunque no fuese el autor material. Un visionado del documental *Capturing the Friedmans* (Andrew Jarecki, dir., 2003) puede ser interesante bajo esta perspectiva. Dos de los protagonistas del documental –padre e hijo– son acusados de pederastia, y ante las condiciones adversas en las que se va a celebrar el juicio deciden declararse culpables pese a considerarse inocentes. Aunque tras ver el documental uno se siente inclinado a pensar que no hay pruebas definitivas ni a favor ni en contra de los hechos que se les imputan, la realidad es que ambos acabaron en prisión, y aunque el hijo quedó en libertad tras cumplir su condena, poca gente estaría dispuesta a contratarle como cuidador infantil.

Estas consideraciones podrían dejar la sensación de que el problema del escepticismo está plenamente vigente en el caso de las personas. Sin embargo las negociaciones narrativas no se realizan a ciegas. Existen factores que las hacen inclinarse en uno u otro sentido de tal forma que, aunque siempre es posible la equivocación, el margen de error se ve disminuido. Para explicarlo podemos recurrir a otros dos conceptos creados en el ámbito de la estética de la recepción: el de «pacto ficcional» –desarrollado por Umberto Eco¹⁶– y el de «horizonte de expectativas» –desarrollado por Hans Robert Jauss–.

El concepto de «pacto ficcional» parte del hecho de que cuando nos enfrentamos a un texto de ficción normalmente parece que suspendemos el juicio sobre la veracidad de lo que se nos está contando. La mayoría de los lectores no creemos que realmente existan los sucesos que se nos cuentan en una novela, pero mientras leemos parece que fingimos creerlo. Sin

16 Umberto Eco no pertenece en realidad a esta corriente teórica, pero sus estudios tienen gran afinidad con ella.

embargo esto no significa que suspendamos *totalmente* nuestro juicio. Aunque la mayoría de nosotros no se preocupa de investigar si existió realmente un cierto personaje literario, no por ello estamos dispuestos a aceptar cualquier cosa que le suceda (Eco, 1994, p. 87). El lector establece con el texto lo que Eco denomina un «pacto ficcional», que marca un límite entre las cosas que se van a considerar verosímiles y las que no. Este pacto lo realiza de forma distinta cada lector individual bajo lo que Jauss denomina su propio «horizonte de expectativas» (Jauss, 1975, pp. 76-78). Al enfrentarse a un texto el lector siempre tiene una serie de expectativas previas al respecto. Puede haber leído otras obras del mismo autor, o del mismo género, puede haber recabado las opiniones de sus amigos, puede estar influido por la publicidad que se le haya hecho al libro... Todos estos condicionantes influyen en lo que el lector espera encontrar o no en el libro y en su experiencia de lectura, interviniendo entre otras cosas en el establecimiento del pacto ficcional. Esto es, en las que cosas que el lector va a aceptar como verosímiles y las que se van a considerar una tomadura de pelo. En este sentido la mayoría de los lectores de *La metamorfosis* de Kafka no tienen ningún reparo en aceptar que un hombre pueda transformarse en un insecto gigante, pues es algo aceptable en una novela de género fantástico. Sin embargo si alguien se acercase a *La metamorfosis* creyendo que se trata de una novela realista encontraría ese mismo suceso totalmente inaceptable, por lo que al romperse el pacto ficcional que había establecido con la obra probablemente abandonase la lectura.

De forma análoga, podemos considerar que en nuestra sociedad las personas establecen con las narraciones de sus congéneres un «pacto de realidad» bajo la influencia de un cierto «horizonte de expectativas». Este horizonte de expectativas está influido por lo que las personas consideran que es esperable de unas narraciones que pertenecen a un cierto tipo de género narrativo: la autobiografía.¹⁷ Pero no sólo por el género, sino también por el conocimiento previo que tienen de esa persona, tanto el adquirido por su propia interacción con ella como el derivado de las narraciones que otros les han contado sobre la misma. De este modo podemos interpretar que cuando se hace necesario efectuar una negociación narrativa para concretizar una indeterminación provocada por versiones contrapuestas, la negociación se guía por el pacto de realidad. Si en un juicio alguien asegura no haber estado en el lugar del crimen en el momento en que se produjo, el pacto de realidad indica que tiene que haber alguna prueba que lo confirme. Y en caso de no haberla, tiene que

17 Para profundizar en esta concepción social del género literario de nuevo puede recurrirse a las ideas de Bajtin, como hace por ejemplo John Shotter (1993).

haber una buena explicación de por qué no la hay. Si la defensa del acusado consiste en asegurar que no pudo cometer el crimen porque en el momento en el que se produjo estaba en el cine, debe haber alguna prueba que lo sitúe allí: el resguardo de una entrada, una grabación de una cámara de seguridad, el testimonio de otros testigos... Si por el contrario su coartada es que no existen pruebas de dónde se encontraba en ese momento porque fue abducido por unos extraterrestres que le llevaron a su nave espacial para preguntarle cuál es el color del cielo visto desde la Tierra, su testimonio será considerado absurdo. Y si insiste en realizar declaraciones que son a todas luces falsas o improbables, finalmente se le considerará como un mentiroso o un demente, modificándose así el horizonte de expectativas que el jurado tiene con respecto a su versión de la historia y perdiendo cada vez más credibilidad. De este modo podemos ver también que el papel de las narraciones en primera persona en una negociación narrativa está determinado por el horizonte de expectativas del resto de implicados, por lo que es variable tanto de una persona a otra como a lo largo del tiempo.

6.2.3. Diferencias entre personas y personajes

El último problema que queda por resolver es el problema de la interpretación. De todo lo que hemos venido diciendo se deduce que las diferencias entre las narraciones que nos constituyen como Yo es o personas y las que constituyen a los personajes literarios no están en el nivel textual, sino en el interpretativo. No se trata de que los personajes literarios tengan características textuales que los Yo es y las personas no tienen, sino del tipo de interpretación que hacemos cuando nos enfrentamos a unos y a otros. El problema consiste entonces en saber qué tipo de interpretaciones es lícito hacer sobre los personajes literarios pero no sobre las personas para asegurarse de que las comparaciones que se hagan entre ambos tipos de entidades estén justificadas.

Como vimos al presentar las críticas de Christman y Lamarque, éstas se basaban en la suposición de que las narraciones literarias tienen unas características formales que las diferencian de un modo absoluto de otro tipo de narraciones. En aquel momento señalamos que dos de estas características eran el poseer una unidad teleológica y una unidad temática, aunque Lamarque refiere algunas más que no consideré necesario reproducir. Y como ya adelanté entonces, en realidad estas características dependen en gran medida de la labor hermenéutica, de tal modo que diferentes intérpretes pueden encontrarlas de forma distinta. Sin

embargo es necesario realizar alguna precisión al respecto, pues podría parecer que esto implica que cada intérprete puede otorgar características a las narraciones de forma arbitraria. Pero que sea posible que distintos intérpretes encuentren características distintas en un mismo texto no implica la existencia de arbitrariedad.

Desde la estética de la recepción se defiende que es el lector el que dota de sentido a los textos. Pero evidentemente esto no significa que les pueda atribuir un sentido cualquiera con independencia de su contenido, pues por ejemplo hay textos a los que ningún intérprete, por imaginativo que sea, podrá dotar de unidad temática, mientras que existen otros a los que difícilmente podría negársela. Como dijimos en los anteriores epígrafes, en los textos existen lugares de indeterminación respecto a los cuales el lector tiene una cierta libertad interpretativa. Pero esto puede suceder únicamente porque en contraposición a ellos existen lugares que están determinados. Si *La metamorfosis* empieza diciéndonos que un hombre se ha transformado en un insecto gigante, es que eso es lo que ha pasado. Las causas del suceso están indeterminadas, pero el resultado no. En este sentido cada texto posee unas características formales que lo hacen más o menos proclive a determinadas interpretaciones al tiempo que vetan otras. Adoptando otra vez conceptos de la estética de la recepción, podemos decir que cada texto tiene su propio «horizonte de expectativas» sobre cómo puede ser interpretado, pues está dirigido a un «lector implícito» del que espera ciertas actitudes (Iser, 1976, pp. 55-70). Una interpretación adecuada es aquella en la que el horizonte de expectativas del lector se encuentra con el del texto, produciéndose así una fusión de horizontes (Jauss, 1975, pp. 76-78).

Si tenemos en cuenta estas restricciones que los textos imponen sobre las interpretaciones que se puede hacer de ellos, podemos decir que lo que convierte en literario a un texto es efectivamente la interpretación que se haga del mismo, pero que dicha interpretación sólo será posible si el texto previamente posee la potencialidad de ser interpretado de ese modo. Sin embargo tener en potencia no es lo mismo que tener en acto. En el horizonte de expectativas de un texto nunca se encuentra una única interpretación, sino un abanico más o menos amplio que puede incluir al mismo tiempo lecturas que lo consideren más o menos literario. Es decir, distintos intérpretes pueden atribuir a una misma narración en distinta medida características que les harían considerarla literaria. En este sentido debemos aceptar que las narraciones literarias no se distinguen de una forma absoluta de las no-literarias, sino que la literariedad es una cuestión de grado. Para mostrar que esto es así puede ser de utilidad volver a reformular la crítica de Lamarque a la comparación entre personas y

personajes prestando atención a las características que atribuye a las obras literarias.

Señala Lamarque que existen dos perspectivas desde las que se puede acceder al contenido de una narración. Una perspectiva interna al mundo que se describe en el texto y una externa. La perspectiva interna consiste en imaginar que los personajes que aparecen en la narración son como personas reales que actúan en un mundo más o menos parecido al nuestro. La perspectiva externa por el contrario no se centra en el mundo que la narración nos hace imaginar sino en el modo (lingüístico) en el que se presenta (Lamarque, 2007, pp. 119-120). Su crítica por tanto consiste en señalar que aunque si interpretamos a los personajes desde la perspectiva interna estos hasta cierto punto se pueden comparar con las personas, reducir los personajes a su dimensión imaginativa supone obviar todo lo que es literario en ellos, esto es, el modo en el que están presentados.¹⁸

Atendiendo por tanto a la perspectiva externa, Lamarque señala que para las narraciones literarias rigen una serie de principios que las personas no cumplen (Lamarque, 2007, pp. 120-128). Además de los principios (1) Teleológico y (2) Temático de los que hablamos en el primer apartado, y (3) el Principio de Identidad de los Personajes que se refiere a lo que denominé el problema de la explicitud, estos son: (4) el Principio de Opacidad, que señala que en las obras literarias nuestra atención se dirige al modo en que se presentan los sucesos y personajes; y (5) el Principio de Funcionalidad, que establece que en las obras literarias siempre es pertinente preguntar por la función estética de cualquier detalle.

Lamarque parece asumir que las obras literarias, por sus propias características formales, cumplen todos estos principios. Sin embargo posiblemente piensa así porque cuando se refiere a «obras literarias» está pensando en piezas de corte clásico que además podrían considerarse como obras maestras de la historia de la literatura. No en vano los ejemplos que utiliza proceden de las novelas de Gustave Flaubert, Charles Dickens o Jane Austen. Si por el contrario concebimos la literatura como

18 La respuesta que Schechtman dio a Lamarque, y que vimos en el primer apartado del capítulo, adquiere todo su sentido en este contexto, pues identifica la perspectiva interna con la perspectiva del personaje y la externa con la del crítico, acusando a Lamarque de adoptar exclusivamente esta segunda con respecto a las obras literarias y la primera con respecto a las personas, cuando las personas tendrían derecho a adoptar tanto la perspectiva del crítico como la del autor, además de la del personaje, con respecto a sí mismas. Sin embargo Schechtman no presenta argumentos contundentes a su favor, por lo que para demostrar que efectivamente las personas pueden adoptar la perspectiva interna o del crítico es necesario continuar de forma independiente a su respuesta.

un fenómeno más amplio se hace evidente que muchas obras literarias no cumplen con estos principios de un modo absoluto, sino sólo hasta cierto punto.¹⁹ Podemos pensar en este sentido en las novelas de Philip K. Dick, uno de los grandes escritores de ciencia-ficción a cuyas obras sin embargo sería exagerado pedirles la misma pulcritud formal e intencionalidad estética con respecto a todos los detalles que a las obras de Flaubert. Por poner otro ejemplo, suponer que una novela como *Peter y Wendy* de James M. Barrie posee un único tema supone mutilarla, pues su interés precisamente radica en que según en qué aspectos nos detengamos encontraremos un tema rector distinto –la importancia de la fantasía en la niñez, el miedo a la muerte, la relación entre la creación artística y la maternidad, la represión de los instintos del hombre en el seno de la sociedad victoriana...– que nos permitirá hacer una interpretación diferente de la obra. Suponer que sólo contiene un único tema e interpretarla en ese sentido nos lleva inevitablemente a una simplificación grotesca como la que hizo Disney en su adaptación al cine de la novela (*vid.* Muñoz Corcuera, 2010b, 2012a).

Pero igual que las narraciones literarias no cumplen necesariamente todos los principios que señala Lamarque, tampoco las narraciones que nos constituyen a nosotros son totalmente ajenas a ellos. Sólo alguien que ignore por completo la dimensión política de nuestras narraciones puede asegurar que el principio de opacidad no rige en nuestro caso porque los modos de presentación no son tan importantes como el contenido presentado. Por otro lado puede que no haya un tema unificador de nuestra vida al completo, pero sin duda hay muchas personas para las cuales gran parte de su vida gira –quizá de modo insano– en torno a una única obsesión. Por último, sin duda no deberíamos exigirles una función estética a todos los aspectos de nuestra vida, pero muchos de ellos pueden tenerlo. Por ponerme a mí mismo de nuevo como ejemplo, una buena cantidad de mis actos de seducción están guiados más por su valor estético que por su utilidad a la hora de alcanzar su supuesto objetivo. Al fin y al cabo el cine ha influido enormemente en nuestra forma de comprender las relaciones humanas y la mayoría de los hombres –si no todos– han deseado parecerse a Humphrey Bogart en alguna ocasión.²⁰

19 De hecho incluso es posible que tampoco las obras maestras en las que está pensando Lamarque los cumplan, aunque prefiero no discutir este aspecto.

20 En este sentido son de gran interés los trabajos de Theodore Sarbin, quien considera que las narraciones míticas y literarias constituyen un modelo de referencia fundamental a la hora de comprender nuestras propias vidas y elaborar las narraciones que constituyen nuestra identidad (*vid.* Sarbin, 1997).

Lo que estos ejemplos pretenden ilustrar es el hecho de que no existen narraciones estrictamente literarias o no-literarias, sino únicamente textos cuya interpretación según unos enfoques es más fructífera que según otros. Sin lugar a dudas interpretar la narración que constituye la vida de una persona atendiendo a su valor estético no dará lugar a resultados tan satisfactorios como la interpretación desde esa perspectiva de una novela, pero esto no significa que no sea posible hacerlo. Del mismo modo que interpretar la narración que constituye a un personaje desde un punto de vista social o psicológico es sin duda menos interesante que interpretar a una persona, aunque existen corrientes de crítica literaria como la sociocrítica y la psicocrítica que se enfocan específicamente en estos aspectos. Si renunciamos a considerar la literariedad como algo que distingue de un modo absoluto unas narraciones de otras podemos ver que estamos justificados para acercarnos a las narraciones que nos constituyen desde cualquier actitud interpretativa. La única consideración que deberíamos tener en cuenta es que quizá la postura que estemos adoptando no vaya a ser la más interesante.

Con una especial sensibilidad hacia el valor de la literatura, el escritor mexicano Jorge Volpi señala que la evolución convirtió a nuestro cerebro en una máquina de predecir el futuro. La literatura habría surgido en este contexto como una herramienta especialmente poderosa que nos ayudaría a predecir el comportamiento de las personas.²¹ A partir de los estudios de Giacomo Rizzolatti, Corrado Sinigaglia o Marco Iacoboni sobre las neuronas espejo, Volpi hace notar que cuando vemos a otra persona realizar una acción, nuestro cerebro reproduce dicha acción como si fuésemos nosotros los que la estamos haciendo. Las neuronas espejo serían así la base neuronal de la empatía, el mecanismo mediante el cual nos ponemos en el lugar del otro para comprender sus acciones presentes y poder predecir las futuras (*vid.* Iacoboni, 2008; Rizzolatti y Sinigaglia, 2006). Lo interesante es que las neuronas espejo no sólo se activan cuando

21 Volpi no es original en este aspecto, pues su aportación al tema –aunque valiosa– es únicamente de divulgación. En este sentido, como señala Garrido Domínguez, ya la intención de Iser con sus teorías era precisamente perseguir «el enraizamiento antropológico de la ficción y su reconocimiento como recurso básico para ensanchar nuestro campo experiencial y la comprensión de nosotros mismos» (Garrido Domínguez, 2011, p. 156). Pero no sólo él, sino que también la poética cognitiva hace uso de las teorías evolutivas para señalar el importante papel de la ficción como recurso para comprender mejor sucesos reales y «refinar nuestra capacidad para procesar información social relevante» (Garrido Domínguez, 2011, p. 172). Especialmente interesante a este respecto puede ser consultar los trabajos de Lisa Zunshine, para quien la ficción nos ayuda a desarrollar nuestra teoría de la mente (*vid.* Zunshine, 2006).

vemos a una persona real hacer una acción, sino también cuando nos imaginamos a un personaje ficticio realizándola (Iacoboni, 2008, p. 14). La literatura, o la ficción en general, sería como un campo de pruebas mediante el cual nos imaginamos personas y situaciones que no han sucedido para saber qué sentiríamos si dichas situaciones llegasen a suceder (Volpi, 2011, p. 22). La función primigenia de la literatura estaría vinculada entonces a lo que Lamarque denomina la perspectiva interna. A la inmersión del lector en el mundo de ficción de tal modo que hace *como si* éste fuese real. Desde esta perspectiva las comparaciones entre personas y personajes son algo muy normal que se da desde los mismos orígenes de la literatura. No porque las personas nos entendamos a nosotras mismas como si fuésemos personajes. Sino porque entendemos a los personajes como si fuesen personas.²²

22 No podemos dejar de señalar que esta es precisamente una de las líneas de investigación que la estética de la recepción desarrolla en la actualidad (*vid.* Robinson, 2010).

CONCLUSIONES

Como se anunciaba en la introducción, el debate sobre la identidad personal en el seno de la filosofía analítica contemporánea se puede considerar que está entrando en una tercera etapa que denomino «nuevo planteamiento narrativo» (NPN). El rasgo característico de este enfoque es la consideración de que somos al mismo tiempo tres entidades distintas –seres humanos, Yoes y personas– cuyas propiedades se habían confundido y mezclado en las fases anteriores de la discusión impidiendo así que se diese una respuesta al problema. El objetivo principal de estas páginas ha sido desarrollar este nuevo planteamiento, aún apenas sugerido por otros autores, para ofrecer una solución consistente a la cuestión de la identidad personal que tuviese en cuenta todos los aspectos de la misma. Esto es, que diese una explicación adecuada de (1) qué es lo que somos y cómo encajan entre sí las dimensiones física, fenoménica y social de nuestra existencia, (2) en qué consiste nuestra persistencia a lo largo del tiempo teniendo en cuenta la gran cantidad de cambios que experimentamos y (3) qué es lo que nos debe importar cuando nos preocupamos por nuestra supervivencia y por la responsabilidad moral. Por otro lado, y teniendo en cuenta las críticas lanzadas contra la anterior etapa del debate con respecto a las comparaciones que se realizaban en su seno entre lo que somos y los personajes literarios, el segundo objetivo ha sido defender (4) que no sólo es posible establecer analogías entre las narraciones que nos constituyen a nosotros y a los personajes de ficción, sino también muy productivo.

En cuanto al primer punto, comencé mostrando los motivos que llevaron a los primeros teóricos del NPN a decir que somos seres humanos, Yoes y personas. Tal y como vimos en el primer capítulo, el planteamiento objetivo del problema de la identidad personal (PO) partía

del presupuesto de que podían ignorarse tanto la dimensión subjetiva de la experiencia como nuestra existencia social, lo que provocó que sus partidarios llegasen a conclusiones inaceptables. Por su parte el planteamiento narrativo clásico (PNC) intentó corregir este error, pero se encontró con la imposibilidad de hacer encajar entre sí estos dos aspectos de nuestra existencia, pues para atender a las demandas que sobre nuestra identidad realizaba una debía desatender las que realizaba la otra. Para tratar de solucionar este problema, los teóricos del NPN –especialmente Marya Schechtman (2007)– sugirieron entonces que tal vez debería postularse la existencia de tres entidades, cada una de las cuales podría dar cuenta de una faceta de lo que somos. El concepto de «ser humano» serviría para atender a nuestra existencia en tanto entidades físicas, el de «Yo» a nuestra dimensión fenomenológica y el de «persona» a nuestra vida social.

Partiendo de la idea de que la diferenciación que se hace en el NPN entre estas tres entidades era el movimiento adecuado para buscar una solución al problema de la identidad personal, y tomando en consideración nuestras intuiciones acerca de lo que nos debe importar cuando nos preocupamos por nuestra supervivencia o por la responsabilidad moral, mostré que el marco ontológico-epistemológico más apropiado para desarrollar el nuevo planteamiento narrativo era una postura intermedia entre el realismo débil de Daniel Dennett (1991b) y el anti-realismo de Nelson Goodman (1978). A partir de ese marco mostré que las distintas posturas interpretativas que podemos adoptar ante la REALIDAD dan lugar a la postulación de la existencia de mundos diferentes e irreducibles entre sí, y que el motivo por el que los teóricos de la identidad personal encontraban tan problemático el atribuirnos al mismo tiempo propiedades físicas, fenoménicas y sociales era que estos tres tipos de propiedades emanan respectivamente de tres actitudes interpretativas distintas. Desarrollando esta idea ofrecí unas definiciones de los conceptos de «ser humano», «Yo» y «persona» que parten de estas tres posturas y que se refieren por tanto a lo que somos en cada uno de los mundos postulados por ellas.

En primer lugar describí aquello que son los seres humanos a partir de la misma actitud interpretativa –que podemos denominar «actitud de diseño»– que utilizamos para comprender el resto de objetos físicos. De este modo aseguré que la descripción más adecuada del concepto de «ser humano» es aquella que ofrece la biología, y que considera que somos un tipo de animal con unas ciertas características. En segundo lugar elaboré una definición de «Yo» desde la actitud intencional definida por Dennett que tiene en cuenta nuestra existencia como sujetos de experiencias fenomenológicas y como agentes auto-conscientes. En consecuencia

afirmé que un Yo es una entidad mental que posee la capacidad de auto-comprenderse en clave narrativa, constituyéndose a sí misma gracias a dicha capacidad de auto-interpretación. Por último definí a las personas desde lo que denominé la «actitud colectiva», que consiste en suponer que un conjunto de sujetos poseen una serie de creencias, deseos e intenciones comunes, y definirles tanto a ellos mismos como al entorno que les rodea en función de estos patrones de intencionalidad colectiva que les hemos adscrito. En este sentido argumenté que las personas son entidades sociales constituidas mediante las creencias colectivas –articuladas en forma de narraciones– que existen sobre ellas en la sociedad en la que viven. Denominé a esta postura «tesis social de la Narratividad».

En cuanto al segundo punto que pretendía aclarar para dar una solución al problema –en qué consiste nuestra identidad a lo largo del tiempo–, dado que en el NPN se asegura que somos tres entidades distintas evidentemente tuve que ofrecer una triple respuesta. Primero argumenté que la identidad de los seres humanos consiste en la continuidad de su cuerpo, al igual que sucede con el resto de objetos físicos y con los animales. Dado que nuestro cuerpo es susceptible de cambios traumáticos –por ejemplo debido a trasplantes de órganos–, la continuidad corporal es una cuestión de grado, pudiendo existir en mayor o menor medida. En este sentido mostré que nuestra identidad en tanto seres humanos es una relación indeterminada, pues siempre es posible que el grado de continuidad que existe entre los cuerpos de dos seres humanos en dos momentos temporales diferentes no sea ni lo suficientemente grande como para aceptar su identidad ni lo suficientemente escaso como para negarla.

Tras ello defendí que la identidad de los Yoes consiste en la continuidad de su conciencia, siendo la narración que cada Yo hace sobre sí mismo para constituirse el elemento capaz de unificar ésta salvando las interrupciones que pudiesen afectarla. De este modo vimos que pueden existir Yoes cuya duración en el tiempo sea apenas la de una experiencia aislada y Yoes cuya narración se prolongue mucho más allá de la vida de un ser humano, sobreviviendo por ejemplo gracias a un cerebro dentro de una cubeta. Teniendo en cuenta que los Yoes se constituyen a sí mismos mediante una narración, mostré que el único requisito para poder decir que alguien continúa siendo el mismo Yo que era en un momento anterior es que él mismo experimente su existencia como continuándose en el tiempo, poseyendo un «acceso empático» a sus propios estados mentales pasados y elaborando una narración congruente con esas sensaciones subjetivas (Schechtman, 2001). Como es obvio, el acceso empático es también una cuestión de grado, pues un Yo puede identificarse en mayor o menor medida con el Yo que experimentó un suceso pasado. En consecuencia

argumenté que la narración que unifica la vida de un Yo puede tener unos límites borrosos, por lo que su identidad es indeterminada, no estando claro en muchas ocasiones si alguien continúa siendo o no el mismo Yo.

Finalmente defendí que la identidad de las personas consiste en la continuidad de las creencias colectivas –o de las narraciones que les dan forma– que las constituyen. Dado que las narraciones que existen sobre cualquier persona en una sociedad son múltiples y a menudo contradictorias, mostré que su identidad se establece mediante una *negociación narrativa* entre todas las posturas implicadas. En esta negociación se deben encontrar siempre en equilibrio la narración que la propia persona ofrece de su vida y las narraciones que sobre ella tienen los demás, pues en principio ninguno de estos dos extremos se debe imponer de tal modo que anule el otro (*vid.* Broncano, 2013; Cruz, 2005). Por otro lado, a pesar de que las narraciones que existen sobre una persona pueden ser contradictorias y por tanto su identidad puede parecer una cuestión de grado, argumenté que el tipo de negociación narrativa que se da en nuestra sociedad no sólo nos permite, sino que también nos mueve a concretizar las ambigüedades presentes siempre que sea necesario, de tal modo que la identidad de las personas es determinada.

En lo que respecta al tercer punto, el principal argumento en el que me basé para demostrar que para desarrollar el NPN había que hacerlo en un marco anti-realista fue, como señalaba anteriormente, que sólo así podríamos ser fieles a nuestras intuiciones acerca de qué es lo que nos debe importar cuando nos preocupamos por nuestra supervivencia y por la responsabilidad moral. Pues como argüí en el tercer capítulo, podemos identificarnos en la misma medida con las tres entidades que somos y preocuparnos en consecuencia por su supervivencia en igual grado. Dado que de las definiciones que ofrecí de los conceptos de «ser humano», «Yo» y «persona» se deriva además que sus condiciones de supervivencia son independientes entre sí, esto significa que podemos alegrarnos por sobrevivir como cualquiera de las tres entidades aunque muramos como alguna de las otras dos, lo cual es difícil de justificar en un marco que no sea el que he expuesto en estas páginas. Por otro lado, teniendo en cuenta que la responsabilidad moral está vinculada a nuestra existencia en sociedad, defendí que lo que nos preocupa o debe preocupar cuando nos preguntamos por este tipo de cuestiones éticas son las personas.

Finalmente, el último punto al que dediqué mi atención fue a la defensa de las similitudes entre las narraciones que constituyen a personajes literarios, Yoes y personas. A este respecto mostré que las críticas lanzadas contra las comparaciones que se hacían entre estas entidades en el planteamiento narrativo clásico se basaban en una

ambigüedad presente en los trabajos de los narrativistas entre dos posibles formas de definir lo que es una narración: una centrada en los rasgos formales de los textos y otra basada en el modo cognitivo utilizado para su interpretación. Sin embargo, pese a que era necesario aclarar dicha ambigüedad para resolver asimismo una serie de problemas asociados a ésta, el grueso de las críticas partió del supuesto de que el sentido relevante de «narración» para las teorías de la identidad debía ser el formalista. En respuesta defendí que si en su lugar utilizamos el sentido cognitivo de «narración» para definirnos tanto a nosotros como a los personajes literarios podemos solventar todos los problemas señalados por las críticas. En este sentido mostré que puede ser de gran ayuda recurrir a la estética de la recepción, una corriente de teoría literaria que se centra en el estudio de la forma en que los textos son interpretados por los lectores. Los conceptos elaborados por esta corriente teórica, como los de «horizonte de expectativas», «lugares de indeterminación» o «proceso de concretización» mostraron un enorme potencial para explicar las características de las narraciones que nos constituyen como Yoes y como personas y su parecido con las que hacen lo propio con los personajes literarios.

ANEXOS

Anexo I

Sobre la relación de identidad

El principio de identidad en lógica es una de las tres leyes clásicas junto al principio de no contradicción y el principio del tercero excluido. Se trata de una relación de equivalencia relativamente sencilla que establece que si dos entidades A y B son idénticas, entonces sólo hay una única entidad que es al mismo tiempo A y B. «A» y «B» son distintos nombres que se refieren a la misma cosa. Por ejemplo Fósforo y Héspero son idénticos entre sí, pues en realidad son un único planeta –Venus– referido con distintos nombres según se vislumbre en la mañana o en la tarde. La relación de identidad posee tres propiedades fundamentales. Es *reflexiva* –un objeto A es idéntico a si mismo–, *simétrica* –si A es idéntico a B, entonces B es idéntico a A– y *transitiva* –si A es idéntico a B, y B es idéntico a C, entonces A es idéntico a C–.

Además de estas tres propiedades básicas se desprende como evidente el principio de la *indiscernibilidad de los idénticos*. El hecho de que si dos objetos son idénticos, entonces comparten todas sus propiedades. Si Fósforo es Héspero, entonces si el primero es un planeta el segundo también lo es. Esta propiedad está sujeta a cierta controversia, pues hay cierto tipo de propiedades que Fósforo y Héspero podrían no compartir, como las creencias que los antiguos griegos poseían sobre ellos. Antes de descubrir que se trataba de una sola «estrella errante», los griegos pensaban que eran dos cuerpos distintos y los personificaban en dos figuras mitológicas diferentes. Sin embargo se puede defender que las creencias de los griegos sobre Héspero y Fósforo no son una propiedad del planeta al que ambos términos se refieren, sino de los propios griegos que poseían las creencias, por lo que podemos mantener el principio de la indiscernibilidad de los idénticos.

Más discutible es el principio inverso a éste, el de la *identidad de los indiscernibles*, que dice que si dos objetos poseen todas sus propiedades en común, entonces son idénticos. Sin embargo dos monedas del mismo cuño pueden compartir todas sus propiedades –forma, tamaño, peso...– y sin embargo ser dos monedas distintas. Para resolver esta contradicción podemos distinguir entre dos tipos de propiedades (cualitativas y relacionales) que dan lugar a dos tipos de identidad: *cualitativa* y *numérica*. Dos monedas del mismo cuño son cualitativamente idénticas si poseen todas sus propiedades cualitativas en común. Sin embargo son numéricamente no-idénticas, pues no comparten sus propiedades relacionales. Es decir, no son una y la misma moneda. De este modo el principio de la identidad de los indiscernibles es aplicable siempre y cuando no confundamos los dos tipos de propiedades.¹

Cuando hablamos de identidad personal, lo que nos preocupa es la identidad numérica de aquello que somos. Mis propiedades cualitativas pueden cambiar con el tiempo –puedo quedarme calvo, aprender a hablar noruego, obtener el título de doctor...– y sin embargo seguir siendo el mismo. Por otro lado alguien podría crear un clon mío que compartiese todas mis características físicas y psicológicas, pero yo no sería ese clon.² Esto es así porque cuando nos preocupamos por nuestra identidad nos preocupan lo que Marya Schechtman llama las «cuatro características» básicas de la existencia de las personas: la supervivencia, el interés por nosotros mismos, la compensación futura y la responsabilidad moral (Schechtman, 1996, p. 2). Yo sobrevivo aunque mis propiedades cualitativas puedan cambiar, mientras que no estoy dispuesto a sacrificar mi bienestar en la actualidad a cambio de que un clon mío sea feliz mañana. Nos preocupamos por nosotros mismos de un modo que no nos preocupamos por los demás –piénsese en la diferencia entre que nos digan que mañana le darán una paliza a una persona que se nos parece mucho y que nos digan que esa persona seremos nosotros–. Y finalmente esperamos

1 En realidad existen más argumentos en contra del principio de la identidad de los indiscernibles, como por ejemplo el presentado por Max Black a partir de la idea de un universo hipotético en el que sólo existiesen dos esferas de hierro con las mismas características. Si este universo existiese, las dos esferas no sólo compartirían todas sus propiedades cualitativas, sino que también compartirían la propiedad relacional de estar situadas a una cierta distancia la una de la otra (*vid.* Black, 1952). En cualquier caso para el tema que nos ocupa no es realmente importante este aspecto, por lo que no me detendré más en él.

2 Aunque un clon mío podría arrebatarme mi identidad como persona si se diesen las circunstancias apropiadas. Para más información véase lo dicho sobre las personas en los capítulos tres y cinco.

que la persona que vaya a la cárcel sea exactamente la misma que cometió el asesinato y no sólo alguien con el mismo aspecto.

Sin embargo, a pesar de que lo que nos preocupa es nuestra identidad numérica, es importante hacer notar que algunos autores consideran que nuestras propiedades cualitativas podrían afectar a esta consideración, ya que determinadas clases de cambio psicológico parecerían destruir nuestra identidad (Parfit, 1984, pp. 375-376). Así por ejemplo si sufriese una amnesia retrógrada completa habría quien consideraría que ya no sería la misma persona que era antes, a pesar de que el cambio sería únicamente cualitativo. Esto es así porque las propiedades psicológicas parecen tener una facilidad mayor que las propiedades físicas para explicar la relación entre la identidad y las «cuatro características» (Schechtman, 1996, pp. 14-18). Pero como trataré de mostrar en el cuerpo del trabajo éste es un problema suscitado por la confusión que existe entre los conceptos de «ser humano», «Yo» y «persona». La identidad de los Yoes —que no la de las personas ni la de los seres humanos— efectivamente puede destruirse si se producen cambios psicológicos, pero esto es debido a que en el marco ontológico en el que éstos se definen no se considera que las propiedades mentales sean cualitativas.

En cualquier caso, estas dudas sobre nuestra existencia con respecto a determinadas situaciones nos hacen introducir una nueva distinción dentro de las relaciones de identidad: pueden ser *determinadas* o *indeterminadas*. Una relación de identidad es indeterminada cuando admite gradaciones y/o preguntas vacías. Por el contrario una relación de identidad es determinada si es del tipo todo o nada y es siempre verificable. Las relaciones de identidad cualitativa admiten gradaciones y son siempre indeterminadas, pues un objeto puede parecerse más o menos a otro. Por su parte las relaciones de identidad numérica, aunque en principio podría parecer que son determinadas —una roca es la misma roca o no lo es, no hay posiciones intermedias—, en realidad pueden ser de ambos tipos (Sanfélix, 1994, pp. 257-258). El famoso ejemplo del barco de Teseo nos puede servir de ilustración.

Después de un tiempo navegando con su barco, Teseo descubre que la nave se ha deteriorado y necesita una reparación, por lo que lleva el buque a un dique seco y comienza a sustituir las partes más dañadas por otras nuevas. Pero según avanzan las reparaciones Teseo cae en la cuenta de que los daños eran más graves de los que había previsto, por lo que un mayor número de piezas deben ser sustituidas. Finalmente, tras un año de reparaciones, todas las piezas del barco han sido reemplazadas y Teseo puede salir a navegar de nuevo. Sin embargo, sin que Teseo lo supiese, uno de sus rivales ha recogido las piezas antiguas del navío según eran

desechadas y las ha vuelto a ensamblar, de forma que ahora existen dos buques que pueden considerarse el barco de Teseo.

La situación involucra a tres barcos. El antiguo barco de Teseo («Antiguo»), el nuevo barco de Teseo («Nuevo») y el barco reconstruido por su rival («Rival»). Las tres naves son cualitativamente bastante parecidas, aunque Rival es completamente idéntica con Antiguo, mientras que Nuevo sólo lo es en menor medida. La cuestión de la identidad cualitativa entre los barcos no presenta problemas. Sin embargo cabe preguntarse cuál de los dos buques que existen tras la reparación, si es que lo es alguno, es idéntico numéricamente con el original. Si el rival de Teseo no hubiese construido un barco, podríamos pensar que Nuevo es numéricamente idéntico con Antiguo, pues esa es la forma normal en que hablamos de identidad entre objetos inanimados. Sin embargo teniendo en cuenta que Rival es cualitativamente idéntico con Antiguo nos sentimos inclinados a pensar que también debe serlo numéricamente. Tenemos buenos motivos para pensar que ambos barcos son idénticos al barco original, sin embargo la relación de identidad es necesariamente una relación uno-uno. Nuevo no es idéntico con Rival, por lo que, teniendo en cuenta que una de las propiedades de la relación de identidad es la transitividad, no es posible que los dos sean a su vez idénticos con Antiguo. A pesar de que tenemos todos los datos acerca de la situación, nos sentimos incapaces de dar una respuesta. En este sentido podemos darnos cuenta de que la pregunta «¿cuál de los dos barcos que existen tras la reparación es idéntico numéricamente con el antiguo barco de Teseo?» es una pregunta vacía, y las distintas respuestas –Nuevo, Rival, ambos o ninguno– no son realmente distintas opciones, sino diferentes descripciones de un mismo hecho (Parfit, 1984, p. 394). No existe una respuesta fija a la pregunta por la identidad del barco de Teseo pues no hay ningún hecho físico adicional que nos permitiese decidir la cuestión. La identidad del barco de Teseo sencillamente es indeterminada.

No obstante podemos darnos cuenta de que la determinación o indeterminación de la identidad de un objeto es relativa a un marco conceptual dado. Siempre que nos encontramos con una pregunta vacía podemos establecer un nuevo criterio o convención que nos permita dar una respuesta. Podríamos decir, por ejemplo, que el verdadero barco de Teseo debe estar formado por las piezas originales, resolviendo así que Rival y Antiguo son el mismo barco, mientras que Nuevo es un buque distinto. Por supuesto eso no implica nada acerca de los hechos físicos. Al dar una respuesta de este tipo al caso no estamos añadiendo nada a Rival para convertirlo en idéntico con Antiguo, ni estamos descubriendo una

propiedad esencial del mismo que antes desconocíamos. Se trata de una decisión convencional.

Volviendo de nuevo sobre el tema de la identidad personal, prácticamente todo el mundo posee una fuerte intuición a favor de que es una relación determinada. Pensamos que en todo caso la pregunta «¿habrá alguien vivo mañana que seré yo?» debe tener una respuesta del tipo sí o no. Sin embargo Parfit demostró que si somos reduccionistas y pensamos que únicamente estamos constituidos por nuestro cuerpo y por los estados mentales asociados al mismo entonces debemos aceptar que nuestras creencias acerca de lo que somos están equivocadas y que nuestra identidad es indeterminada igual que la de otros objetos físicos. De lo contrario deberíamos aceptar que las personas bien existen de un modo independiente a sus cuerpos –son una entidad puramente mental, espiritual o compuesta de una sustancia física aún desconocida– o bien no existen de un modo independiente de sus cuerpos, pero su identidad consiste en un hecho adicional desconocido, por lo que tenemos que renunciar al reduccionismo (Parfit, 1984, p. 388). En respuesta a Parfit, espero mostrar que la creencia de que la identidad de las personas es determinada pese a que sólo estamos constituidos por nuestro cuerpo y por los estados mentales asociados al mismo procede también de la confusión entre los conceptos de «ser humano», «Yo» y «persona».

Finalmente podemos distinguir dos tipos de preguntas o juicios relativos a la identidad de un objeto. Podemos preguntarnos en primer lugar acerca de cómo es posible *identificarlo* entre otros objetos parecidos. Este tipo de pregunta es fácil de responder atendiendo a las propiedades cualitativas del objeto en cuestión. Así por ejemplo las personas se pueden identificar de forma bastante fiable por sus huellas dactilares, o por su ADN. Sin embargo lo que en filosofía se conoce como el problema de la identidad personal no se refiere a este tipo de preguntas, de las que ya se encarga la policía, sino de las preguntas acerca de qué *constituye* la identidad de un objeto. Lo que nos interesa es saber cuáles son las condiciones necesarias y suficientes para garantizar su identidad. En este sentido, a pesar de la fiabilidad que puedan tener las pruebas de identificación, las personas no *son* su ADN y a nadie se le ocurriría zanjar el debate filosófico de la identidad personal con una apelación a la intransferibilidad metafísica de las huellas dactilares (Rodríguez González, 2003, p. 40). A pesar de que ni siquiera dos gemelos idénticos tendrían ambas características en común –las huellas dactilares no dependen únicamente de factores genéticos, sino también de factores ambientales intrauterinos– siempre cabría la posibilidad de una coincidencia, o incluso de un engaño realizado mediante una complicada cirugía. Por otro lado

una persona puede perder ambas manos sin que por ello su identidad se vea afectada.

Este tipo de preguntas sobre la constitución de la identidad cobran sentido dentro del problema más específico de la identidad a través del tiempo, cuyos orígenes se remontan a Heráclito y a su afirmación de que es imposible bañarse dos veces en el mismo río, pues éste se encuentra en constante cambio. En general aceptamos que hay cambios que no afectan a la identidad de los objetos y que un río sigue siendo el mismo aunque el agua en su cauce sea diferente. La pregunta filosófica por la identidad es por tanto acerca de cuáles son las condiciones necesarias y suficientes para que un objeto siga siendo el mismo a pesar de los cambios que pueda sufrir. No obstante cada tipo de objeto puede diferir en las condiciones necesarias y suficientes para garantizar su identidad (Perry, 2002, p. 70). En el caso de los objetos físicos –por ejemplo un río– se suele considerar que esta relación es la continuidad espacio-temporal. Un río en el momento t_2 sigue siendo el mismo que era en el momento t_1 si existe una continuidad espacio-temporal entre ambos. Es decir, si hay una línea a través del espacio-tiempo que une ambos estados del río y en cada punto de esa línea existe un río cuya existencia es al menos parcialmente debida a la existencia del río en el punto inmediatamente anterior.³ Sin embargo en el caso de las personas existen en general tres tipos de posturas. Están aquellos que defienden que las personas están constituidas únicamente por sus cuerpos, por lo que las condiciones para su identidad son las mismas que las del resto de objetos físicos. Por otro lado están los que defienden que las personas están constituidas sólo por sus características mentales, siendo la continuidad de su conciencia la relación que estamos buscando. Finalmente se encuentran quienes defienden que las personas están constituidas por una sustancia mental o espiritual independiente del cuerpo, y es a la continuidad de esta sustancia a la que debemos atender. Una vez más, espero demostrar en este trabajo que las discrepancias sobre las condiciones que garantizan nuestra identidad proceden principalmente

3 Parfit señala que es discutible que la identidad de *todos* los objetos físicos dependa de su continuidad espacio-temporal. Un río –o una piedra– puede conservar su identidad a pesar de los cambios que sufra. Sin embargo si un artista pinta un autorretrato, y sobre él pinta un retrato de otra persona no podemos decir que ambas obras sean estados de un mismo cuadro. A pesar de que existe una continuidad espacio-temporal el retrato original ha sido destruido y el nuevo retrato es una obra diferente (Parfit, 1984, pp. 377-378). Lo que Parfit no tiene en cuenta es que las obras de arte no son únicamente objetos físicos, sino también y sobre todo entidades sociales. En cualquier caso no me detendré más en ello.

de la confusión entre los términos «ser humano», «Yo» y «persona», pues a pesar de las apariencias no todos los autores están hablando de lo mismo cuando hablan de personas.

Anexo II

Introduction

‘I know who I am.’

A statement that, no doubt, many of the philosophers who will appear through this work would subscribe. Perhaps without noticing the literary load that the judgment has to almost any Spanish-language reader.

The most famous knight in history – victim of the first of his stumbles even before his adventures had started – is being escorted back to his home by a humble peasant, neighbour of his. Without knowing where he is, Don Quixote mistakes the farmer with the Marquis of Mantua, one of the characters in a romance, and himself with his nephew Baldwin. The newly appointed marquis, still not entirely convinced of the insanity of the gentleman, tries to make him understand:

‘Senor – sinner that I am! – cannot your worship see that I am not Don Rodrigo de Narvaez nor the Marquis of Mantua, but Pedro Alonso your neighbour, and that your worship is neither Baldwin nor Abindarraez, but the worthy gentleman Senor Quixada?’ (Cervantes, 1605, p. 122)

In that precise moment, when reality comes to impose itself, we can imagine our hero, wounded but proud, lifting his head to face it:

‘I know who I am,’ replied Don Quixote, ‘and I know that I may be not only those I have named, but all the Twelve Peers of France and even all the Nine Worthies, since my achievements surpass all that they have done all together and each of them on his own account.’ (Cervantes, 1605, pp. 122-123)

‘I know who I am’ has been the leitmotif of the discussion on personal identity in analytic philosophy since the mid-twentieth century. And from

it they have tried to answer some of the most important questions that philosophy can contemplate; since issues such as those regarding our survival or moral responsibility hang from the problem of personal identity.⁴ How to know if I will be alive tomorrow if I do not know who I am? Of which actions will I have to respond? Of whom future should I worry? No doubt their research would have been in a better track if they had given more space from the beginning to the contributions literature could make to it.

But we can not complain. Narrative identity theories made their way into analytic philosophy decades ago. And the comparisons between the narratives that (they say) constitute us as persons and those that constitute literary characters came along with them. Comparisons that have been called into question in recent dates and that I want to defend here along with a personal development of what I call the new narrative approach (NNA).

The current debate on personal identity has its origins in the mid-twentieth century when authors such as Bernard Williams (1957), Sydney Shoemaker (1959), Derek Parfit (1971), John Perry (1972), and David Lewis (1976) picked up the argument made by Locke in his *Essay Concerning Human Understanding*. For Locke ‘person’ is a term – which in this context should be understood as synonymous with ‘self’ – that refers to a mental entity distinct from the human being in whose body is embodied.⁵ His identity is defined in terms of the continuity of a single consciousness, memory being the element capable of unifying it through everyday interruptions caused by sleep or short periods of oblivion (Locke, 1690, p. 318-319). What is interesting in Locke’s position is that he considers that when we care about our survival or moral responsibility – what Marya Schechtman calls the *four features*⁶ of personal identity

4 For more information regarding the problem of personal identity see Appendix I, which should be read as a large footnote to this introduction.

5 It is important to note that the debate on the problem of personal identity in the terms that I will find interesting here is totally alien to dualism. In fact although Locke admits the existence of the soul, he considers that personal identity does not depend on its continuity (Locke, 1690, p. 319). In this sense the term ‘embodied’, which will appear quite frequently throughout this work, does not presuppose in any case the prior existence of an entity which is separable from the body and which is then embodied in it. The solutions to the problem of personal identity proposed from dualism will not be considered.

6 Schechtman effectively points out four features: the prudential concern for our own future, the future compensation for the present situations, survival and moral responsibility. However I think that the first two are linked enough with the problem of survival as not to consider them independently.

(Schechtman, 1996, p. 136-162) – what matter to us is our existence as persons, not as human beings. In this sense if the consciousness of a human being is interrupted so that he does not remember a crime he committed in the past, it should not be attributed to him, since strictly speaking it would have been committed by a person other than the one that currently inhabits the body of the human being in question.

Locke's theory, despite its problems, has great virtues. However, it was understood by his critics as a simplistic theory based on memory continuity, what had two consequences. The first is that it was proved to be patently false since the eighteenth century, receiving harsh attacks that were not answered in centuries (*vid.* Butler, 1736; Reid, 1785). The second is that the criterion proposed by Locke to determine the identity of a person based on memory continuity was understood by his readers in the twentieth century as an objective criterion that could be opposed to other objective criterion: bodily continuity. So, ignoring the subjective dimension of consciousness whose continuity was what defined for Locke the existence of persons (*cf.* McDowell, 1997), the debate that occurred since the mid-twentieth century to about the 1980's was dominated by what I call the objective approach (OA).

All the participants in the debate that occurred within the OA accepted Locke's premise that the term 'person' should be used to refer to the entity that should matter to us when we worry about the four features. However, the two opposing sides argued over which was that entity. Supporters of physical criteria – Williams was his most significant defense – argued that 'person' was synonymous with 'human being,' so that our identity should be defined in terms of the continuity of our body as it happens with the rest of the animals. But supporters of psychological criteria – among which we can highlight Shoemaker, Lewis, Parfit and Perry – certainly represented the dominant position. For these authors, heirs to the Lockean position, the term 'person' should be considered synonymous with 'self' and therefore refer to what we are as mental entities. In spite of that, distancing themselves from the English philosopher, they also argued that persons should be defined in terms of psychological (instead of consciousness) continuity, as a result of the objectivist interpretation that they made of Locke's theory. His argument was based on the fact that, although normally the physical and psychological continuity coincide, if we imagine a hypothetical case in which they diverged – for example due to a brain transplant – what we care about when we worry about the four features seems to be the psychological continuity and not the physical one.

The debate within the OA reached its peak (and in a sense its end) with the publication of Parfit's *Reasons and Persons* in 1984. In his book the

philosopher from Oxford defended a version of the neo-Lockean position, showing both its virtues and its limitations. For though he maintains along with the rest of the psychological continuity supporters that what should matter to us when we worry about our survival is our mental life, he also argues convincingly that it is not necessarily related to personal identity. The reason is that, mainly due to its connection with moral responsibility and our existence in society, personal identity must be determinate. That is, there should always be a yes or no answer to the question about a person's identity. Since psychological continuity is a matter of degree and it can also take a branched form – two future persons could both be at the same time psychologically continuous with a single past person – it can not be an equivalent relationship to that of personal identity, even if normally both relationships happen together. In this sense Parfit argues that what should matter to me when I care about my survival shouldn't be that I am the one who survives, but it would be enough that there is someone who is similar enough to me. That is, there is someone psychologically continuous with me even though he doesn't have my identity. In this sense we can understand the mantra which Parfit was repeating since time ago: personal identity is not what matters (Parfit, 1984, p. 458).

The extravagance of this statement, and especially its inevitability within the framework of the OA, made that from the eighties a new approach started to break into the debate about personal identity. It is what I call the classic narrative approach (CNA). Among its supporters we can highlight Alasdair MacIntyre (1981), Jerome Bruner (1987), Daniel Dennett (1988), Charles Taylor (1989), Paul Ricoeur (1990), and Marya Schechtman (1996). It can be interpreted that the theories of these authors try to reformulate the psychological continuity of the objective approach in terms of the continuity of the narrative we tell about ourselves, so in a sense they inherited the assumptions of the neo-Lockean side (*vid.* Slors, 1998). But they do so by departing from two criticisms to the OA that will serve to them as a basis to develop the main thesis that underpin the CNA. The first criticism relates to the fact that one of the most important properties of our existence is that we have consciousness, a radically subjective phenomenon that can not be reduced to objective facts such as psychological continuity. The second states that the moral responsibility issues which are linked to the concept of personal identity doesn't have a meaning outside the social space in which they arise, so that the primacy that the neo-Lockeans gave within the OA to the own person's opinion when deciding on their identity incapacitated them to capture the social dimension of our lives. Basically both criticisms had a common goal: to

note that the objectivist framework of the former approach was unable to account for both aspects of our existence, since it considered them to be reducible to a physical or psychological objective basis (in fact Parfit denominates his own position as ‘reductionist’).

To solve the first problem and attend to the subjective dimension of consciousness CNA theorists defended what following Galen Strawson may be called the psychological Narrativity thesis (*vid.* Strawson, 2004). This thesis says that persons subjectively experience their lives in narrative form, so that the psychological continuity posited by the neo-Lockeans in the OA as the criterion of identity should be replaced by the continuity of the narrative with which we understand our lives. Meanwhile they tried to solve the second problem – the inability to capture the social dimension of our existence if persons are defined solely from what they think about themselves – with the ethical Narrativity thesis, which states that understanding in a single narrative one’s life as a whole is necessary to have a good life and achieve full personhood. The idea was that by understanding themselves in this way persons would take into account the context in which they develop their actions and would therefore give space for the social aspect of their existence. To ensure that this was so, some authors as MacIntyre and Ricoeur postulated that the own person’s narrative should enter into dialogue with the narratives that others tell about them, so that their identity was the result of this interaction (MacIntyre, 1981, p. 263; Ricoeur, 1990, p. 161-162). However, the dominant position in this regard was Schechtman’s, who simply postulated the need for our narrative about ourselves to submit, at least to some extent, to the real events in which the person was involved (Schechtman, 1996, pp. 119-130). Thus for most of CNA supporters our identity was constituted only by our own narrative about ourselves, which caused the ethical Narrativity thesis to be unable to completely solve the problem which gave rise to it. But despite this, narrative identity theories were received with great enthusiasm, and since the eighties until recently they were accepted almost without opposition.

However, despite the initial enthusiasm, in recent years the CNA has been subject to numerous criticisms that have finally led to its withdrawal. Simplifying to the maximum the sense of these objections, it is possible to group them into two blocks. The first, which we can consider led by Strawson’s article ‘Against Narrativity’ to which I referred in the previous paragraph (Strawson, 2004), focuses on the fact that if our identity is constituted by a single narrative, it’s impossible that it meets at the same time all the facets of our existence. The second, whose main exponent is Peter Lamarque (2004, 2007), accuses narrativists of having used the term

‘narrative’ ambiguously, assuming that all narratives are literary and thus abusing in an unjustified way of the comparisons between the narratives that supposedly constitute persons and that which constitute literary characters.

This second block of criticisms is undoubtedly of great importance, and some of its aspects that have not yet received a reply will occupy my attention in the last chapter. However it is the first set of criticisms that has received most of the attention and the one that has eventually led to the collapse of the CNA. And that’s because, as Schechtman showed when trying to respond to Strawson, what his criticisms reveal is that there is an underlying tension in the classic narrative approach between the demands that the subjective and social dimensions of our existence make on the story that constitutes us (Schechtman, 2007, p. 168-169). If we try to fit our narrative as much as possible to our subjective experience of the world, we end in having to give up any significant sense of ‘narrative’ that allows us to give an account of our social existence. In fact, some people say they do not experience their lives as narratives at all – Strawson himself claims to be an example of this – so that if the CNA were true we would have no choice but, counter-intuitively, denying them personhood. If on the other hand we try to make our life-narrative complex enough to account for its social dimension, we are forced to give up to explain the way some persons subjectively perceive their own lives. Thus, the CNA faces the impossibility of proposing a single narrative entity which simultaneously possesses the necessary properties to account for the phenomenological and social dimensions of our existence.

Faced with this impasse, the solution offered by Schechtman constitutes the end of the CNA and the basis on which to develop the new narrative approach (NNA): to finish with the synonymy between ‘self’ and ‘person.’ Schechtman thinks that the best alternative to avoid the problems revealed by Strawson’s criticisms is to assume that (1) we are not only the two entities that had been postulated in the debate from the beginning – mental selves and physical human beings – and (2) that the term ‘person’ should not be used to refer to that of the two that matter to us when we care about the four features. Instead we should consider that the term ‘person’ has its own reference, other than those of ‘self’ and ‘human being,’ and that this reference is an entity somehow not sufficiently specified related to the social sphere. Thus the NNA is characterized by the defense that we are at the same time three distinct entities – human beings, selves and persons – and that at least one of them – the person – is constituted in narrative form.

However, once we have decided that the term ‘person’ is no longer to be used to designate the entity that matter to us when we care about the four features but to refer to an entity whose existence had not been contemplated before, it is not clear which of the three entities that we are should matter to us when we care about our survival or moral responsibility. In fact the few authors who have contributed to the NNA until today – especially Schechtman (2007, 2011), Dan Zahavi (2007) and Richard Menary (2008) – have conflicting opinions about it. In any case it could not have been otherwise, since due to its recent emergence nobody has been interested yet in discussing the matter in order to develop further this new approach. In fact there isn’t consensus still even on whether selves are constituted narratively or not, or on which is exactly the relationship of persons with the social sphere.

As I announced at the beginning of this introduction, the aim of these pages will be precisely to develop the NNA. This will take as its starting point the solution Schechtman offered in response to the first block of criticisms received by the CNA. However, I will distance myself from her at two points. The first is that, since one of the reasons to differentiate between selves and persons in the NNA is the need to give an appropriate account of the social dimension of our existence, in order to develop my concept of ‘person’ I will recur – rather than to the characterization of the concept Schechtman gave – to the suggestions made by MacIntyre and Ricoeur on the role that the narratives told by others about ourselves take in the constitution of our identity. The second is that I will defend that our intuitions invite us to consider that the four features of personal identity Schechtman was talking about are not applicable as a group to one of the three entities that are distinguished in the NNA. As I will try to show, when we care about our survival we care about the future of the human being, the self, and the person we are to the same extent. But we will see that when we speak of moral responsibility we only care about persons, which is not surprising considering that the concept of ‘person’ is defined in the NNA specifically to meet the demands of our social life.

Once I have defended these two points, I will show that both are incompatible with a materialist framework, so I will stand for its abandonment. Evidently this will be one of the most controversial claims of my work, since until recent dates materialism, in its various forms, has virtually been the only acceptable position in the field of analytic philosophy of mind. However we must keep in mind that the difficulty of explaining mental qualities within a standard materialist framework – the famous explanatory gap that seems to force us to accept an eliminative position (Levine, 1983) – has caused in recent decades that some

alternative positions as Antonio Damasio's emergentism (Damasio, 2003, p. 302) or Strawson (1994) and David Chalmers' panpsychism (1996) gain strength. Regardless of their value, what these positions reveal is the philosophical need to find an alternative to standard materialism. But as I will try to show, none of them has enough consistency to be considered an appropriate framework within which to solve neither the problem of consciousness nor to justify our intuitions about that when we care about our survival we are concerned to the same extent for the three entities we are. Being so, I hope that my bet on abandoning a materialist framework will be seen not only as a plausible alternative but also desirable.

In any case, my intention is not to defend some kind of dualism as an alternative to materialism. Such a project would certainly be doomed. Instead I will present a proposal that will resume a line of thought that can be considered Wittgensteinian: the mind-body problem is a false problem which is derived from the misapplication of some concepts – those concerning mental properties – in a framework that is completely alien to them – materialism – so there is no possible physical explanation of consciousness. The reason for this incapability however is not ontological, as the emergentists and panpsychists seem to suggest, but epistemological.

The main problem to defend my position will be the belief that materialism is not only a satisfactory explanation of reality, but the only possible, as its suitability to it is not only epistemological, but also ontological. That is, reality is not only *explained* by materialism, but it *is* as materialism suppose it to be, so if the concepts related to mental properties are not applicable in a materialistic framework, then they do not exist at all. For this reason first I will argue that no scientific discourse can follow realistic criteria when developing its conceptual framework, but only pragmatic criteria, since its adequacy to reality will always be untestable (*vid.* Quine, 1948). As a result of this I will show that depending on the interpretative stance we adopt toward this objective and uncognizable reality – hereafter REALITY – we can build different valid explanations – that is, useful – that are incompatible with each other (*vid.* Dennett, 1991b). Thus each of these explanations will presuppose the existence of a number of different entities which will also be irreducible to a common framework, so we will be able to say that each of them postulates the existence of a different subjective world or reality (*vid.* Goodman, 1978). Of course none of these worlds will have an objective REAL existence, but they will be as much real as we can get to know. REALITY is uncognizable, so that none of the statements that we can make will never have an objective ontological value, but solely one relative to our subjective experience of the world.

Once I have defended my position with regard to the ontological and epistemological problems of scientific knowledge I will show that we can solve not only the problem of consciousness, but we can also justify our intuitions about what should matter to us in our survival and thus solve the problem of personal identity. To do so we will only have to assume that the concepts of ‘self,’ ‘person,’ and ‘human being’ belong to three different worlds which are the result of using three different interpretative stances: Human beings are physical entities defined from what we can call a *design stance*; the selves are phenomenological entities defined from an *intentional stance*; and for its part persons are social entities belonging to a world described from what I will call a *collective stance*. In this way we will be able to see that, although their survival conditions and identity characteristics are independent because they emanate from the world they belong to, we can identify ourselves to the same degree with the three entities.

One aspect that is worth noting is that the definition of persons within the NNA will be the most original. First because the interpretative attitude from which I will define them – the collective stance – does not come directly from Dennett’s theory as it happens with the design and intentional stances, although it is based on it (*vid.* Dennett, 1991b). And second, because as I will be inspired to define them by MacIntyre and Ricoeur’s positions on the role of the narratives of others, the result will be that our identity as persons will be constituted through the collective beliefs about us which exist in our society, beliefs that will be articulated through a *narrative negotiation* between our own life story and the stories that others tell about us. These considerations will give rise to what I will call the *social Narrativity thesis*, one of the main contributions of this work and that could be considered as an improvement of the ethical Narrativity thesis.

Coming back to the way in which the personal identity problem is solved in the NNA, once we have a proper definition of ‘human being,’ ‘self,’ and ‘person’ I will show that precisely the lack of precision in these concepts was the origin of the problem. So for example the participants in the OA debate, as they did not properly distinguish between the characteristics of each entity, tried to ascribe a social property – the determinacy of identity – now to human beings now to selves, whose identities are necessarily a matter of degree. On the other hand most of the theorists of CNA, as well as the neo-Lockeans of the OA, considered that our identity was defined only from the first person perspective – the ‘I know who I am’ of which I was talking about at the beginning of this introduction – regardless of the fact that although if the nobleman Alonso

Quijana loses his mind and thinks to be a character in a novel of chivalry it can cause him to be no longer the same self that he was previously, he will still remain the same human being he was before he lost his head, and not even all the speeches in the world could convince his neighbour Pedro Alonso – nor, by the way, any person in their right mind – that he should give him the honors which the one who has decided to be renamed as Don Quixote claims for himself.

Once my development of the NNA had been explained to this point I hope the reasoning presented will be sufficient to consider the problem of personal identity solved. However, the second block of criticisms made against the CNA, referred to the alleged abuse of the comparison between persons and characters based on an ambiguous use of the term ‘narrative,’ can also be applied to the NNA. Being so, it will be necessary to address these criticisms to give greater strength to my thesis.

The solution will depart from the recognition of the *ambiguity problem* in the CNA, as well as that of other four derivative problems: the *triviality problem*, the *explicitness problem*, the *skepticism problem*, and the *interpretation problem*. When trying to prove that our identity is constituted by a narrative, CNA theorists oscillate from a cognitive sense of ‘narrative’ – that emanates from our way of understanding human actions – and a formalist sense – linked to the formal features that a text should have to be a literary narrative – so that they only demonstrate that our thinking is narrative in the first sense but they give an unjustified step and assume that it must also be so in the second one. However I will show that the criticisms made by Lamarque start from the assumption that the relevant sense of ‘narrative’ for CNA theorists is the formalist, when actually it is the cognitive. If we assume that this is so, then both the ambiguity problem and the other four derivative problems can be solved if when making comparisons between persons and literary characters we consider that we should not define the latter in formalists terms, but from a framework closer to the cognitive sense of ‘narrative,’ such as that used by reader-response criticism (*vid.* Mayoral, 1987; Warning, 1975). Starting then from the way this literary theory conceptualizes the narratives which appear in the novels I will show that comparisons between selves, persons and characters are not only possible, but also extremely productive.

Taking the above into consideration, the present work can be divided into two parts:

The first, which is essentially a review of the previous approaches, is devoted to the historical evolution of the debate on personal identity in analytic philosophy until the appearance of NNA. It consists of two chapters. The first focuses on the analysis of the OA from its origin until

its abandonment and the second does the same with the CNA, showing the reasons that cause the appearance of the NNA.

The second part, which consists on the remaining four chapters, is more argumentative and aims to answer the criticisms received by the CNA and to develop the NNA in order to solve the problem of personal identity. The third chapter focuses on the defense of (1) that when we care about moral responsibility we only care about persons, (2) that when we care for our survival we care to the same extent about the three entities that we are, and (3) that the latter is unjustifiable in a materialist framework. The fourth offers an anti-realist alternative to standard materialism that would allow us to solve the problem of consciousness and to justify the fact that we care to the same extent for the survival of human beings, selves and persons. In the fifth chapter I define these three entities, taking into account the anti-realist framework presented in the previous chapter, after which I show how the problem of personal identity – understood as the question about what and who we are, what our persistence over time consists in and what the entity or entities that should matter to us when we worry about the four features is – can be solved without troubles in the NNA. Finally in the sixth chapter I discuss the charges leveled against the CNA in regard of their use of the term ‘narrative,’ answering them so that it becomes evident that comparisons between selves, persons and characters in the NNA are both possible and useful.

As I said at the very beginning, the sentence ‘I know who I am’ seems to have been the leitmotif of the discussion on the problem of personal identity in analytic philosophy since its origins. This assumption has led to arguments based on what each of the participants in the debate had the impression to be, as I noted in this introduction and we will see later. This work aims to escape from this trend by offering a broader conception of what it is what we mean when we say ‘I,’ which includes what each of us believe to be but is not limited to it. I hope that the arguments presented here will convince the reader and will open new perspectives with respect to the problem of personal identity.

Anexo III

Conclusions

As it was announced in the introduction, it can be considered that the debate about personal identity in contemporary analytic philosophy is entering a third stage which I call ‘new narrative approach’ (NPN). The characteristic feature of this approach is the consideration that we are at the same time three distinct entities – human beings, selves, and persons – whose properties had been confused and mixed in the previous stages of the discussion thus making impossible to give an answer to the problem. The main purpose of these pages has been to develop this new approach, barely suggested by other authors, to provide a consistent solution to the question of personal identity that takes into account all of its aspects. That is, one that gives an adequate explanation of (1) what we are and how do the physical, phenomenal and social dimensions of our existence fit together, (2) in what consists our persistence over time taking into account the many changes we experience and (3) what should matter to us when we worry about our survival and moral responsibility. On the other hand, considering the criticisms leveled against the previous stage of the debate regarding the comparisons that were made within it between what we are and literary characters, the second aim has been to defend (4) that it is not only possible to draw analogies between the narratives that constitute ourselves and the ones that constitute fictional characters, but also very productive.

Regarding the first point, I began by showing the reasons that led the first NNA theorists to say that we are human beings, selves and persons. As we saw in the first chapter, the objective approach to the problem of personal identity (OA) was based on the assumption that both the

subjective dimension of experience and our social existence could be ignored, thus causing its supporters to arrive to unacceptable conclusions. For its part the classic narrative approach (CNA) attempted to correct this mistake, but it was unable to fit together these two aspects of our existence, as to meet the demands about our identity made by one of them it had to ignore the ones made by the other. To try to solve this problem, theorists of NNA – especially Marya Schechtman (2007) – suggested that maybe it should be posited the existence of three entities, each of which could account for a facet of what we are. The concept of ‘human being’ would take care of our existence as physical entities, that of ‘self’ of our phenomenological dimension and that of ‘person’ of our social life.

Departing from the idea that the differentiation proposed within the NNA between these three entities was the right move to find a solution to the problem of personal identity, and taking into consideration our intuitions about what should matter to us when we worry about our survival or moral responsibility, I showed that the most appropriate ontological and epistemological framework to develop the new narrative approach was an intermediate position between Daniel Dennett’s mild-realism (1991b) and Nelson Goodman’s anti-realism (1978). From this framework I showed that the different stances we can adopt to interpret reality lead to the postulation of the existence of different worlds which are irreducible to each other, and that the reason why theorists of personal identity found so problematic to attribute to ourselves at the same time physical, social, and phenomenal properties was that these three types of properties emanate from three different stances respectively. Developing this idea I offered a definition of the concepts of ‘human being,’ ‘self,’ and ‘person’ that depart from these three attitudes and therefore which are related to what we are in each of the worlds postulated by them.

First I described what human beings are from the same interpretative attitude – which we can call ‘design stance’ – we use to understand the rest of physical objects. Thus I assured that the most appropriate description of the concept of ‘human being’ is the one biology offers, which considers that we are a kind of animal with certain characteristics. Secondly I developed a definition of ‘self’ from the intentional stance defined by Dennett which takes into account our existence as both subjects of phenomenological experiences and self-conscious agents. Therefore I said that selves are mental entities that have the ability to understand themselves in narrative form, constituting themselves through such capacity for self-interpretation. Finally I defined persons from what I called the ‘collective stance,’ which consists in postulating that a set of subjects share certain beliefs, desires, and intentions, and defining both

themselves and the environment around them according to these patterns of collective intentionality that we have ascribed to them. In this sense I argued that persons are social entities constituted by the collective beliefs – which are articulated in narratives – that exist about them in the society in which they live. I named this the ‘social Narrativity thesis.’

Regarding the second point I intended to clarify in order to give a solution to the problem – in what consists our identity over time – since the NNA ensures that we are three distinct entities evidently I had to offer a triple response. First I argued that the identity of human beings consists in the continuity of their body, as it happens with the rest of physical objects and animals. Since our body is susceptible to traumatic changes – for example due to an organ transplant – bodily continuity is a matter of degree that can exist to a greater or lesser extent. In this regard I showed that our identity as human beings is an indeterminate relationship, since it is always possible that the degree of continuity between the bodies of two human beings existing in two different times is neither large enough to accept their identity nor low enough to deny it.

After that I argued that the identity of selves consists in the continuity of their consciousness, being the narrative that each self tells to constitute himself the element capable of unifying it thus saving the disruptions that could affect it. In this regard we saw that there may be selves whose time duration is hardly an isolated experience and selves whose narrative extends far beyond the life of a human being, for example by surviving as a brain in a vat. Considering that selves constitute themselves through a narrative, I showed that the only requirement for someone to say that he is still the same self he was at an earlier time is that he experiences its existence as being continued in time, possessing an ‘empathic access’ to their own past mental states and developing a narrative consistent with those subjective sensations (Schechtman, 2001). Obviously, empathic access is also a matter of degree, since a self can identify himself to a greater or lesser extent with the self who experienced a past event. Therefore I argued that the narrative that unifies the life of a self can have fuzzy limits, so that his identity is indeterminate, not being clear in many cases whether someone is still the same self or not.

Finally I defended that the identity of persons consists in the continuity of the collective beliefs that constitute them – or that of the narratives that shape the beliefs –. Since the narratives existing about anyone in a society are multiple and often contradictory, I showed that their identity is established through a *narrative negotiation* between all the involved postures. In this negotiation the narrative that the person offers about his own life and the narratives that the others tell about him must always be in

equilibrium, since in principle none of these extremes should be imposed so that it cancels the other (*vid.* Broncano, 2013, Cruz, 2005). On the other hand, although the narratives existing about a person can be contradictory and therefore his identity may seem a matter of degree, I argued that the kind of narrative negotiation that occurs in our society does not only allow us, but induces us to concretize the present ambiguities whenever necessary, so that the identity of persons is determinate.

Regarding the third point, the main argument in which I relied to show that to develop the NNA it had to be done in an anti-realist framework was, as noted above, that it was the only way we could be faithful to our intuitions about what is what should matter to us when we care about our survival and moral responsibility. As I argued in the third chapter, we can identify ourselves to the same extent with the three entities that we are and consequently be concerned for their survival to the same degree. Given that the fact that their survival conditions are independent of each other is derived from the definitions of the concepts of ‘human being,’ ‘self,’ and ‘person’ that I offered, this means we can be happy to survive as any of the three entities even if we die as one of the other two, which is difficult to justify in a framework other than that I have discussed in these pages. On the other hand, considering that moral responsibility is linked to our existence in society, I argued that persons are what should matter to us when we care for this type of ethical issues.

Finally, the last point to which I turned my attention was the defense of the similarities between the narratives that constitute literary characters, selves and persons. In this regard I showed that the criticisms leveled against the comparisons which were made between these entities in the classic narrative approach were based on an existing ambiguity in the work of the narrativists between two possible ways of defining what a narrative is: one focused on the formal features of the texts and another based on the cognitive mode used for their interpretation. However, despite the need to clarify this ambiguity in order to solve also a number of problems associated with it, the bulk of the criticisms departed from the assumption that the relevant sense of ‘narrative’ for identity theories should be the formalist. In response I argued that if instead we use the cognitive sense of ‘narrative’ to define us as well as literary characters we can solve all the problems pointed out by the criticisms. In this regard I showed that it can be helpful to turn to reader-response criticism, a trend in literary theory that focuses on the study of the way in which texts are interpreted by readers. The concepts developed by this theoretical trend, such as ‘horizon of expectations,’ ‘gaps of indeterminacy,’ or ‘concretization process,’ showed an enormous potential to explain the

features of the narratives that constitute us as selves and persons and their similarities with those which do the same with literary characters.

BIBLIOGRAFÍA

- Bajtín, M. (1961/2002). Para una reelaboración del libro sobre Dostoevski. En *Estética de la creación verbal* (pp. 324-345). Buenos Aires: Siglo XXI.
- . (1975/1989). *Teoría y estética de la novela*. Madrid: Taurus.
- Baker, L. R. (2000). *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Ballard, J. G. (1990/2008). *Fiebre de guerra*. Córdoba: Berenice.
- Benito Vicente, J. Ó. (2003). El problema de la identidad personal en la filosofía analítica. *Daimon: Revista de filosofía*, (28), 67-83.
- Bermúdez, J. L. (1997). Reduction and the Self. *Journal of Consciousness Studies*, 4(5-6), 458-466.
- Black, M. (1952). The Identity of Indiscernibles. *Mind*, 61(242), 153-164.
- Block, N. (1978). Troubles with Functionalism. En C. W. Savage (Ed.), *Minnesota Studies in the Philosophy of Science, IX* (Perception and Cognition: Issues in the Foundation of Psychology), 261-325.
- Bordes Solanas, M. (1998). Filosofía en clave de ciencia ficción: Las personas y sus condiciones de supervivencia en el tiempo. *Teorema*, 17(2), 59-75.
- Broncano, F. (2013). *Sujetos en la niebla: Narrativas sobre la identidad*. Barcelona: Herder.
- Bruner, J. S. (1986/1988). *Realidad mental y mundos posibles: Los actos de la imaginación que dan sentido a la experiencia*. Barcelona: Gedisa.
- . (1987). Life as Narrative. *Social Research*, 54(1), 11-32.
- . (1991). The Narrative Construction of Reality. *Critical Inquiry*, 18(1), 1-21.

- . (1997). A Narrative Model of Self Construction. *Annals of the New York Academy of Sciences*, 818(1), 145-161.
- Butler, J. (1736/2008). Of Personal Identity. En J. Perry (Ed.), *Personal Identity* (pp. 99-106). Berkeley y Los Angeles: University of California Press.
- Carr, D. (1986). Narrative and the Real World: An Argument for Continuity. *History and Theory*, 25(2), 117-131.
- . (1998). Phenomenology and Fiction in Dennett. *International Journal of Philosophical Studies*, 6(3), 331-344.
- Cervantes, M. de. (1605/1987). *Don Quijote de la Mancha. Primera Parte*. (Vol. I). Madrid: Gredos.
- Chalmers, D. J. (1996/1999). *La mente consciente: En busca de una teoría fundamental*. Barcelona: Gedisa.
- . (2010). *The Character of Consciousness*. Oxford: Oxford University Press.
- Christman, J. (2004). Narrative Unity as Condition of Personhood. *Metaphilosophy*, 35(5), 695-713.
- Churchland, P. M. (1984/1992). *Materia y conciencia: Introducción contemporánea a la filosofía de la mente*. Barcelona: Gedisa.
- . (1995). *The Engine of Reason, the Seat of the Soul: A Philosophical Journey Into the Brain*. Cambridge, Mass.: MIT Press.
- Colomina Almiñana, J. J. (2008). Una solución materialista a la corazonada «zombie». *Revista de filosofía*, 33(2), 161-174.
- Conrad, J. (1899/2005). *El corazón de las tinieblas*. Madrid: Cátedra.
- Cosentino, E. (2011). Self in Time and Language. *Consciousness and Cognition*, 20(3), 777-783.
- Cruz, M. (1999). *Hacerse cargo: Sobre responsabilidad e identidad personal*. Barcelona: Paidós.
- . (2005). *Las malas pasadas del pasado: Identidad, responsabilidad, historia*. Barcelona: Anagrama.
- Currie, G. (2010). *Narratives and Narrators: A Philosophy of Stories*. Oxford: Oxford University Press.
- Damasio, A. R. (2003/2005). *En busca de Spinoza: Neurobiología de la emoción y los sentimientos*. Barcelona: Crítica.
- . (2010). *Y el cerebro creó al hombre: ¿Cómo pudo el cerebro generar emociones, sentimientos, ideas y el yo?* Barcelona: Destino.
- Dennett, D. C. (1976/1989). *Condiciones de la cualidad de persona*. México: Universidad Nacional Autónoma de México.
- . (1987/1998). *La actitud intencional*. Barcelona: Gedisa.
- . (1988/1992). The Self as a Center of Narrative Gravity. En F. S. Kessel, P. M. Cole, y D. L. Johnson (Eds.), *Self and Consciousness: Multiple Perspectives* (pp. 103-115). Hillsdale, NJ: Erlbaum.

- . (1991a/1995). *La conciencia explicada: Una teoría interdisciplinar*. Barcelona: Paidós.
- . (1991b). Real Patterns. *The Journal of Philosophy*, 88(1), 27-51.
- . (2001). The Zombic Hunch: Extinction of an Intuition. En A. O'Hear (Ed.), *Philosophy at the New Millenium* (pp. 27-43). Cambridge: Cambridge University Press.
- Diéguez Lucena, A. (1998). *Realismo científico: Una introducción al debate actual en la filosofía de la ciencia*. Málaga: Universidad de Málaga.
- Donato-Rodríguez, X. de (2009). Construction and Worldmaking: the Significance of Nelson Goodman's Pluralism. *Theoria*, 24(3), 213-225.
- Draaisma, D. (1995/1998). *Las metáforas de la memoria: Una historia de la mente*. Madrid: Alianza.
- Eco, U. (1977/1983). *Cómo se hace una tesis*. Barcelona: Gedisa.
- . (1994/1996). *Seis paseos por los bosques narrativos*. Barcelona: Lumen.
- Feyerabend, P. K. (1975/1979). El mito de la ciencia y su papel en la sociedad. *Cuadernos Teorema*, 53, 11-36.
- Frankfurt, H. G. (1976/2006). Identificación y externalidad. En *La importancia de lo que nos preocupa* (pp. 89-103). Buenos Aires: Katz.
- Gagnon, J. H. (1992). The Self, Its Voices, and Their Discord. En C. Ellis y M. G. Flaherty (Eds.), *Investigating Subjectivity: Research on Lived Experience* (pp. 221-243). Newbury Park, CA.: SAGE Publications.
- García Ruiz, P. E. (2009a). Configuraciones del yo: Parfit y Ricoeur sobre la identidad personal. *Analogía filosófica: Revista de filosofía, investigación y difusión*, 23(2), 137-153.
- . (2009b). El nudo del mundo: Subjetividad y ontología de la primera persona. *Eidos*, 10, 194-223.
- Garrido Domínguez, A. (2004). El texto literario a la luz de la hermenéutica. *Signa*, (13), 103-124.
- . (2011). *Narración y ficción: Literatura e invención de mundos*. Madrid y Frankfurt: Iberoamericana y Vervuert.
- Gazzaniga, M. S. (1998/1999). *El pasado de la mente*. Barcelona: Andrés Bello.
- Gergen, K. J. (1991/1992). *El yo saturado: Dilemas de identidad en el mundo contemporáneo*. Barcelona: Paidós.
- Goodman, N. (1978/1990). *Maneras de hacer mundos*. Madrid: Visor.
- . (1989). Just the Facts, Ma'am! En M. Krausz (Ed.), *Relativism: Interpretation and Confrontation* (pp. 80-85). Notre Dame, IN: University of Notre Dame Press.

- Greene, B. (2004/2005). *The Fabric of the Cosmos: Space, Time and the Texture of Reality*. London: Penguin.
- Hofstadter, D. R., y Dennett, D. C. (Eds.). (1981/1983). *El ojo de la mente: Fantasías y reflexiones sobre el yo y el alma*. Buenos Aires: Editorial Sudamericana.
- Hutto, D. D. (2007). The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology. En D. D. Hutto (Ed.), *Narrative and Understanding Persons* (pp. 43-68). Cambridge: Cambridge University Press.
- Iacoboni, M. (2008/2009). *Las neuronas espejo: Empatía, neuropolítica, autismo, imitación o de cómo entendemos a los otros*. Madrid: Katz.
- Ingarden, R. (1968/1989). Concreción y reconstrucción. En R. Warning (Ed.), *Estética de la recepción* (pp. 35-53). Madrid: Visor.
- Iser, W. (1972/1987). El proceso de lectura: enfoque fenomenológico. En J. A. Mayoral (Ed.), *Estética de la recepción*. Madrid: Arco Libros.
- . (1976/1987). *El acto de leer: Teoría del efecto estético*. Madrid: Taurus.
- Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly*, 32(127), 127-136.
- Jauss, H. R. (1975/1987). El lector como instancia de una nueva historia de la literatura. En J. A. Mayoral (Ed.), *Estética de la recepción*. Madrid: Arco Libros.
- Kandel, E. R. (2007). *En busca de la memoria: El nacimiento de una nueva ciencia de la mente*. Buenos Aires: Katz.
- Kermode, F. (1966/1983). *El sentido de un final*. Barcelona: Gedisa.
- Kivy, P. (2010). The Experience of Reading. En G. L. Hagberg y W. Jost (Eds.), *A Companion to the Philosophy of Literature* (pp. 106-119). Sussex: Wiley-Blackwell.
- Korsgaard, C. M. (1989). Personal Identity and the Unity of Agency: A Kantian Response to Parfit. *Philosophy & Public Affairs*, 18(2), 101-132.
- Kripke, S. (1980/1985). *El nombrar y la necesidad*. México: Universidad Nacional Autónoma de México.
- Kristjánsson, K. (2009). Realist Versus Anti-Realist Moral Selves—and the Irrelevance of Narrativism. *Journal for the Theory of Social Behaviour*, 39(2), 167-187.
- Kuhn, T. S. (1962/2004). *La estructura de las revoluciones científicas*. México: Fondo de Cultura Económica.
- Lamarque, P. (2004). On Not Expecting Too Much from Narrative. *Mind & Language*, 19(4), 393-408.

- . (2007). On the Distance between Literary Narratives and Real-Life Narratives. En D. D. Hutto (Ed.), *Narrative and Understanding Persons* (pp. 117-132). Cambridge: Cambridge University Press.
- Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*, 34, 354-361.
- Lewis, D. K. (1976/1983). Survival and Identity. En *Philosophical Papers* (Vol. 1, pp. 55-77). Oxford: Oxford University Press.
- Locke, J. (1690/1982). *Ensayo sobre el entendimiento humano*. México: Fondo de Cultura Económica.
- Lodge, D. (2003/2004). *La conciencia y la novela: Crítica literaria y creación literaria*. Barcelona: Península.
- MacIntyre, A. (1981/2004). *Tras la virtud*. Barcelona: Crítica.
- Mayoral, J. A. (Ed.). (1987). *Estética de la recepción*. Madrid: Arco Libros.
- McCarthy, J. (2007). *Dennett and Ricoeur on the Narrative Self*. Amherst, NY: Humanity Books.
- McCrone, J. (2003). Not-so total recall. *New Scientist*, 178(2393), 26-29.
- McDowell, J. (1997). Reductionism and the First Person. En J. Dancy (Ed.), *Reading Parfit* (pp. 230-250). Oxford: Blackwell.
- McKoon, G., y Ratcliff, R. (1992). Inference During Reading. *Psychological Review*, 99(3), 440-466.
- Menary, R. (2008). Embodied Narratives. *Journal of Consciousness Studies*, 15(6), 63-84.
- Moya, C. J. (2004). *Filosofía de la mente*. Valencia: Universitat de València.
- Muñoz Corcuera, A. (2010a). Memoria, ficción y realidad en Mulholland Drive y Carretera Perdida de David Lynch. *Especulo: Revista de estudios literarios*, (44). Recuperado a partir de <http://www.ucm.es/info/especulo/numero44/dvlynch.html>
- . (2010b). Las curiosas aventuras de Peter Pan en el mundo del celuloide (I). *Anuario de Investigación en Literatura Infantil y Juvenil*, (8), 85-110.
- . (2012a). Reescribiendo Peter Pan: La indefinición de un mito con múltiples originales. *Anuario de Investigación en Literatura Infantil y Juvenil*, (10), 287-325.
- . (2012b). The True Identity of Captain Hook. En A. Muñoz Corcuera y E. T. Di Biase (Eds.), *Barrie, Hook, and Peter Pan: Studies in Contemporary Myth; Estudios sobre un mito contemporáneo* (pp. 66-90). Newcastle: Cambridge Scholars Publishing.
- . (2013). Problemas y aciertos de la teoría del Yo narrativo de Dennett: Aportaciones al debate sobre la identidad personal. *Logos: Anales del seminario de metafísica*, 46.

- Nelson, K. (2003). Narrative and the Emergence of a Consciousness of Self. En G. D. Fireman, T. E. McVay, y O. J. Flanagan (Eds.), *Narrative and Consciousness* (pp. 17-36). New York: Oxford University Press.
- Olson, E. T. (1997). *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press.
- . (2007). *What are We? A Study in Personal Ontology*. New York: Oxford University Press.
- Parfit, D. (1971/1983). *Identidad personal*. México: Universidad Nacional Autónoma de México.
- . (1984/2004). *Razones y personas*. Boadilla del Monte (Madrid): A. Machado Libros.
- . (2004). *Personas, racionalidad y tiempo*. Madrid: Síntesis.
- . (2012). We Are Not Human Beings. *Philosophy*, 87(1), 5-28.
- Penrose, R. (1989/2002). *La mente nueva del emperador: En torno a la cibernética, la mente y las leyes de la física*. Madrid: Mondadori.
- . (1994/1996). *Las sombras de la mente: Hacia una comprensión científica de la consciencia*. Barcelona: Crítica.
- Perry, J. (1972). Can the Self Divide? *The Journal of Philosophy*, 69(16), 463-488.
- . (1975/2008). (Ed.). *Personal Identity*. Berkeley y Los Angeles: University of California Press.
- . (2002). The Two Faces of Identity. En *Identity, Personal Identity and the Self* (pp. 64-83). Indianapolis, IN: Hackett.
- Pinker, S. (1997/2007). *Cómo funciona la mente*. Barcelona: Destino.
- Pirandello, L. (1904/1998). *El difunto Matías Pascal*. Madrid: Cátedra.
- Quine, W. O. (1948/2002). Acerca de lo que hay. En *Desde un punto de vista lógico* (pp. 39-59). Barcelona: Paidós.
- . (1950/2002). Identidad, ostensión e hipóstasis. En *Desde un punto de vista lógico* (pp. 113-130). Barcelona: Paidós.
- . (1968/1986). Relatividad ontológica. En *La relatividad ontológica y otros ensayos* (pp. 43-91). Madrid: Tecnos.
- . (1972). Review of Identity and Individuation, by Milton K. Munitz. *The Journal of Philosophy*, 69(16), 488-497.
- Reid, T. (1785/2008). Of Mr. Locke's Account of Our Personal Identity. En J. Perry (Ed.), *Personal Identity* (pp. 113-118). Berkeley y Los Angeles: University of California Press.
- Ricoeur, P. (1980). Narrative Time. *Critical Inquiry*, 7(1), 169-190.
- . (1983/2004). *Tiempo y narración I: Configuración del tiempo en el relato histórico*. México: Siglo XXI.

- . (1984/2008). *Tiempo y narración II: Configuración del tiempo en el relato de ficción*. México: Siglo XXI.
- . (1985/2009). *Tiempo y narración III: El tiempo narrado*. México: Siglo XXI.
- . (1990/1996). *Sí mismo como otro*. Madrid: Siglo XXI.
- Rizzolatti, G., y Sinigaglia, C. (2006). *Las neuronas espejo: Los mecanismos de la empatía emocional*. Barcelona: Paidós.
- Robinson, J. (2010). Emotion and the Understanding of Narrative. En G. L. Hagberg y W. Jost (Eds.), *A Companion to the Philosophy of Literature* (pp. 71-92). Sussex: Wiley-Blackwell.
- Rodríguez González, M. (2003). *El problema de la identidad personal: Más que fragmentos*. Madrid: Biblioteca Nueva.
- . (2010). La conciencia fenoménica y el límite del Naturalismo. *Convivium: Revista de filosofía*, 23, 173-188.
- Rorty, A. O. (1976). A Literary Postscript: Characters, Persons, Selves, Individuals. En A. O. Rorty (Ed.), *The Identities of Persons* (pp. 301-323). Berkeley and Los Angeles: University of California Press.
- Rudd, A. (2007). In Defence of Narrative. *European Journal of Philosophy*, 17(1), 60-75.
- Sanfélix, V. (1994). Las personas y su identidad. *Logos: Anales del seminario de metafísica*, (28), 257-285.
- Sarbin, T. R. (1997). The Poetics of Identity. *Theory Psychology*, 7(1), 67-82.
- Schaffer, J. (2007). From Nihilism to Monism. *Australasian Journal of Philosophy*, 85(2), 175-191.
- Schechtman, M. (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- . (2001). Empathic Access: The Missing Ingredient in Personal Identity. *Philosophical Explorations*, 4(2), 95-111.
- . (2007). Stories, Lives and Basic Survival: A Refinement and Defense of the Narrative View. En D. D. Hutto (Ed.), *Narrative and Understanding Persons* (pp. 155-178). Cambridge: Cambridge University Press.
- . (2011). The Narrative Self. En S. Gallagher (Ed.), *The Oxford Handbook of the Self* (pp. 394-416). Oxford: Oxford University Press.
- Scheffler, I. (1980). The Wonderful Worlds of Goodman. *Synthese*, 45(2), 201-209.
- Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417-424.
- . (1992/1996). *El redescubrimiento de la mente*. Barcelona: Crítica.
- . (1995/1997). *La construcción de la realidad social*. Barcelona: Paidós.

- Sellars, W. (1963/1971). *Ciencia, percepción y realidad*. Madrid: Tecnos.
- Shoemaker, S. (1959). Personal Identity and Memory. *The Journal of Philosophy*, 56(22), 868-882.
- . (1963). *Self-Knowledge and Self-Identity*. New York: Cornell University Press.
- . (1970). Persons and Their Pasts. *American Philosophical Quarterly*, 7(4), 269-285.
- . (1985/1997). Parfit on Identity. En J. Dancy (Ed.), *Reading Parfit* (pp. 135-148). Oxford: Blackwell.
- Shotter, J. (1993). Becoming Someone: Identity and Belonging. En N. Coupland y J. F. Nussbaum (Eds.), *Discourse and Lifespan Identity* (pp. 5-28). Newbury Park, CA.: SAGE Publications.
- Slors, M. (1998). Two Conceptions of Psychological Continuity. *Philosophical Explorations*, 1(1), 61-80.
- Snowdon, P. (1990). Persons, Animals, and Ourselves. En C. Gill (Ed.), *The Person and the Human Mind: Issues in Ancient and Modern Philosophy* (pp. 578-596). Oxford: Oxford University Press.
- . (2011). The Animal You Are. *The Philosopher's Magazine*, 54(3), 35-43.
- Stokes, P. (2010). What's Missing in Episodic Selfhood? A Kierkegaardian Response to Galen Strawson. *Journal of Consciousness Studies*, 17(1-2), 119-143.
- . (2012). Is Narrative Identity Four-Dimensionalist? *European Journal of Philosophy*, 20(Supplement 1), 86-106.
- Strawson, G. (1994/1997). *La realidad mental*. Barcelona: Prensa Ibérica.
- . (1997). «The Self». *Journal of Consciousness Studies*, 4(5-6), 405-428.
- . (1999). The Self and the SESMET. *Journal of Consciousness Studies*, 6(4), 99-135.
- . (2004/2008). Against Narrativity. En *Real Materialism and Other Essays* (pp. 189-207). Oxford: Oxford University Press.
- . (2006). Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies*, 13(10-11), 3-31.
- . (2007). Why I have no future. *The Philosopher's Magazine*, 38(2), 21-26.
- . (2009). *Selves: An Essay in Revisionary Metaphysics*. New York: Clarendon Press.
- . (2012). We Live Beyond Any Tale That We Happen to Enact. *Harvard Review of Philosophy*, 18, 73-90.
- Strawson, P. F. (1959/1989). *Individuos: Ensayo de metafísica descriptiva*. Madrid: Taurus.

- Taylor, C. (1989/1996). *Fuentes del yo: La construcción de la identidad moderna*. Barcelona: Paidós.
- Teichert, D. (2004). Narrative, Identity, and the Self. *Journal of Consciousness Studies*, 11(10-11), 175-191.
- Todorov, T. (1995/2000). *Los abusos de la memoria*. Barcelona: Paidós.
- Velleman, J. D. (2005/2006). The Self as Narrator. En *Self to Self: Selected Essays* (pp. 203-223). Cambridge: Cambridge University Press.
- Vice, S. (2003). Literature and the Narrative Self. *Philosophy*, 78(303), 93-108.
- Volpi, J. (2011). *Leer la mente: El cerebro y el arte de la ficción*. México D. F.: Alfaguara.
- Walker, M. J. (2012). Neuroscience, Self-Understanding, and Narrative Truth. *AJOB: Neuroscience*, 3(4), 63-74.
- Warning, R. (Ed.). (1975/1989). *Estética de la recepción*. Madrid: Visor.
- White, H. (1980). The Value of Narrativity in the Representation of Reality. *Critical Inquiry*, 7(1), 5-27.
- Wilkes, K. V. (1988). *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.
- Williams, B. (1957). Personal Identity and Individuation. *Proceedings of the Aristotelian Society*, 57, 229-252.
- . (1960). Bodily Continuity and Personal Identity: A Reply. *Analysis*, 21(2), 43-48.
- . (1970). The Self and the Future. *The Philosophical Review*, 79(2), 161-180.
- . (2007). Life as Narrative. *European Journal of Philosophy*, 17(2), 305-314.
- Wilson, R. A. (2005). Persons, Social Agency, and Constitution. *Social Philosophy & Policy*, 22(2), 49-69.
- Zahavi, D. (2007). Self and Other: The Limits of Narrative Understanding. En D. D. Hutto (Ed.), *Narrative and Understanding Persons* (pp. 179-201). Cambridge: Cambridge University Press.
- Zunshine, L. (2006). *Why We Read Fiction: Theory of Mind and the Novel*. Columbus: Ohio State University Press.