UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

Departamento de Economía Aplicada III (Política Económica)

TESIS DOCTORAL

Essays in Popperian epistemology, rationality and macroeconomics

Ensayos de epistemología popperiana, racionalidad y macroeconomía

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Iván Huáscar Ayala García

Director

Alfonso Palacio-Vera

Madrid, 2016

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

Departamento de Economía Aplicada III (Política Económica)

TESIS DOCTORAL

Essays in Popperian Epistemology, Rationality & Macroeconomics

Ensayos en Epistemología Popperiana, Racionalidad y Macroeconomía

Iván Huáscar Ayala García

Director:

Alfonso Palacio-Vera

Index

Resumen

Este conjunto de ensayos tiene como principal objetivo proporcionar una base epistemológica a las ciencias sociales y en especial, a la economía. Nuestro punto de partida y principal premisa es la Teoría Evolucionista Popperiana del Conocimiento y del Aprendizaje (PTKL). No obstante, existe una segunda rama en la obra de Popper que está directamente relacionada con la metodología de las ciencias sociales: el Análisis Situacional (SA) y el Principio de Racionalidad (RP). De estas dos ramas se derivan dos nociones distintas de "racionalidad" en la obra de Popper. El objetivo del presente estudio, por tanto, es doble: (i) estudiar la relación entre PTKL y SA y (ii) estudiar los fundamentos epistemológicos de dos enfoques macroeconómicos rivales a la luz de la PTKL y el RP, a saber, la teoría keynesiana y la llamada "Nueva Síntesis Neoclásica".

El primer ensayo es un análisis crítico de la relación entre PTKL y el SA y las implicaciones que se derivan para las ciencias sociales. Nuestra primera hipótesis es que existe cierta tensión entre el PTKL y el SA cuando su relación se analiza desde el punto de vista de la “racionalidad de los agentes” cuyo comportamiento es recogido en el modelo, aunque la tensión desaparece en gran medida cuando dicha relación se analiza desde el punto de vista de la “racionalidad del modelizador”. Nuestra segunda hipótesis es que el tipo de tensión entre el PTKL y el SA depende de si el modelizador adopta la versión “objetivista” (*RPo*) o “subjetivista” (*RPs*) del *RP*. Nuestra tercera hipótesis se apoya en las ideas de Hayek (1943) sobre los “hechos” de las ciencias sociales, y argumentamos que tal y como Popper y otros autores definen el *RPo,* éste representa un caso extremo pues presupone que el P-S es completamente *independiente* de las expectativas de los agentes. Nuestra cuarta y última hipótesis está estrechamente relacionada con la anterior, y consiste en la idea de que, si aceptamos las ideas de Hayek sobre la naturaleza de los hechos en las ciencias sociales, la estrategia natural para los

científicos sociales es la reconstrucción del P-S tal y como los *agentes la perciben*, más que la reconstrucción de acuerdo a cómo el *científico la percibe*. Por último, y a diferencia de lo que Popper y otros autores sugieren, la diferencia entre el *RPo* y el *RPs* no consiste en que en el primero el modelizador reconstruye el P-S *como realmente es* y, que en el segundo, se reconstruye *como los agentes lo ven*, sino que en el primero el modelizador reconstruye el P-S *como él percibe que el P-S es* mientras que en el segundo lo reconstruye *como él cree que los agentes creen que P-S es*.

En el primero de los ensayos apuntamos que Hume era escéptico acerca de la racionalidad del comportamiento humano porque pensaba que éste estaba basado en la "costumbre y el hábito", es decir, en el supuesto de que el "futuro será como el pasado" a pesar de no existir una "lógica inductiva". Igualmente, en sus escritos tardíos, Keynes presupone que tomamos decisiones utilizando procedimientos inductivos que consisten en adoptar reglas sociales y convenciones en las que confiamos debido a los buenos resultados generados en el pasado. Las reglas y convenciones que Keynes identifica en sus escritos tardíos son, a todos los efectos, equivalentes a la vieja idea de Hume acerca del comportamiento basado en "costumbres y hábitos". Concretamente, la convención más importante es el supuesto de que "el estado presente de las cosas continuará de forma indefinida a no ser que haya razones específicas para esperar un cambio" (Keynes, 1936, p. 152). La tesis principal del ensayo es que la teoría de Keynes es compatible con la PKTL. Además, mostramos que el tipo de metodología implícita en la teoría keynesiana es un caso particular de *RPs*. Respecto al punto anterior, mostramos que la formación convencional de expectativas en la teoría macroeconómica keynesiana puede ser interpretada como el resultado último de un complejo proceso caracterizado por el método de prueba y error donde aquellas convenciones que son percibidas por los agentes como erróneas son reemplazadas en el tiempo por nuevas convenciones que emergen

espontáneamente. Proponemos el concepto de “Convenciones Hegemónicas” (HC) para denotar estas convenciones que son utilizadas por los agentes económicos porque creen que representan una guía válida para la toma de decisiones. En la medida que el proceso por el que las HC emergen y desaparecen es un proceso de prueba y error, establecemos una analogía entre el crecimiento del conocimiento convencional en la teoría keynesiana y el proceso de expansión del conocimiento derivado del PTKL. A continuación, argumentamos que la convención “el presente estado de las cosas continuará de forma indefinida a no ser que haya razones específicas para esperar un cambio” está en el núcleo de la teoría keynesiana y es responsable de dos de las críticas más habituales que ha recibido: (i) el supuesto comportamiento irracional de los agentes y (ii) su aparente nihilismo. En el ensayo mostramos que estas críticas son una consecuencia directa de la adopción por parte de los economistas ortodoxos del *RPo* y que pueden ser debidamente contestadas cuando se demuestra que la teoría keynesiana es, contrariamente a la teoría neoclásica, una versión particular del *RPs*.

El último ensayo analiza la “Hipótesis de Expectativas Racionales” (REH) a la luz de la PTKL. En el ensayo se defiende que la adopción de la REH sesga la teoría macroeconómica y reduce su relevancia práctica al restringir su aplicabilidad a situaciones de “riesgo” (Knight 1971[1921]) o “mundos pequeños” (Savage, 1954). En dichos contextos, la única fuente de falibilidad es la incapacidad de los agentes económicos para anticipar shocks aleatorios “exógenos”. Por el contrario, en el ensayo se señala que la PKTL proporciona a la teoría macroeconómica una sólida fundamentación epistemológica que permite incorporar plenamente la falibilidad humana.

Abstract

These essays represent an attempt to provide an epistemological basis for the social sciences and, especially, for economics. Our departure point is Popper's Evolutionary Theory of Knowledge and Learning (PTKL). However, there is a second strand in Popper´s philosophy that is directly related to the methodology of the social sciences: Popper´s Situational Analysis (*SA*) and the `Rationality Principle´ (*RP*) that lies at its core. It follows that there are two notions of rationality in Popper´s work: (i) the notion that stems from (PTKL) according to which human behaviour is `rational´ if it exhibits `corrigibility´ (i.e., if individuals are willing to eliminate their errors), and (ii) the notion that stems from Popper´s *RP* according to which human behaviour is `rational´ if it is appropriate or adequate the Problem-Situation (P-S) in which actors find themselves. Therefore, the purpose of this study is *two-fold*: (i) to analyze the relation between PTKL and *RP*, and (ii) to analyze the epistemological foundations of two different (and rival) macro-theories (the so-called New Neoclassical Synthesis and Keynes´s macro-theory) in the light of both PTKL and *SA*.

The first essay provides a critical analysis of the relationship between PTKL and SA and its implications for the social sciences. Our first claim is that there is a certain tension between PTKL and *SA* when their relation is analyzed from the standpoint of the `rationality of the agents´ whose behaviour the theoretical model seeks to capture albeit the tension disappears when the relation is analyzed from the standpoint of the `rationality of the theoretician´. Our second claim is that the nature of the tension between PTKL and *SA* depends upon whether the theoretician adopts the `objectivist´ (*RPo*) or the `subjectivist´ (*RPs*) version of *RP*. Our third claim builds upon the ideas of Hayek (1943) about the `facts´ of the social sciences and is that, as presented by Popper and his commentators, *RPo* represents an extreme case based on the presupposition that P-S is

(fully) *independent* of agents´ beliefs. Our fourth claim is closely related to the previous one and consists of the idea that, if Hayek´s ideas are taken on board, it follows that the natural strategy for social scientists is to seek to reconstruct P-S *as agents´ see it* rather than to reconstruct it *as the scientist sees it*. Our last claim is that, unlike what Popper suggests, the difference between *RPo* and *RPs* is not that in the former the theoretician reconstructs P-S *as it actually is* whereas in the latter she does it as *agents see it* but, rather, that in the former she reconstructs P-S *as she sees it* herself whereas in the latter she does it *as she believes that agents actually see it.*

In our second essay we note that Hume was sceptical about the rationality of human behaviour because he believed that the latter was grounded upon `custom and habit´, that is, on the assumption that the `future will resemble the past´ in spite of the inexistence of an `inductive logic´. Now, the view Keynes adopts in his later economic writings consists of the notion that we make decisions by using inductive procedures which essentially amount to adopting the social rules and conventions which are widely believed to have yielded good results in the past. The former are, for our purposes, equivalent to Hume´s old idea that individual behaviour is grounded upon `custom and habit´. In particular, the chief convention of all, according to Keynes, is the assumption that `the existing state of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change´. This assumption plays an essential role in Keynesian macro-theory in that it enables the theoretician to construct models in which: (i) economic agents behave in a `rational´ way in a context of genuine uncertainty, and (ii) testable predictions can *a priori* be generated. The main thesis of this essay is that Keynes´s theory is compatible with PTKL. Further, we also argue that the methodology underlying Keynes´s macro-theory is an instance of the `subjectivist´ version of Popper´s `Rationality Principle´ (*RPs*). Finally, we argue that the convention that `the existing state

of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change´ is at the core of Keynesian macro-theory and is directly responsible for two criticisms the latter has been subject to: (i) the alleged `irrational´ behaviour of Keynesian economic agents, and (ii) the *nihilism* of Keynesian theory. Notwithstanding it, we argue that these criticisms stem from the adoption by mainstream economists of the `objectivist´ version of Popper´s `Rationality Principle´ (*RPo*) and that they can be duly answered by showing that Keynes´s theory is instead an instance of *RPs*.

The third and last essay analyses the `Rational Expectations Hypothesis´ (REH) in the light of PTKL. We argue that the adoption of REH biases macroeconomics and reduces its practical relevance by restricting its applicability to situations of 'risk' or 'small worlds'. In such contexts, the only source of fallibility is the inability of economic agents to anticipate 'exogenous' random shocks. By contrast, we argue that PTKL provides macro-theory with solid epistemological and ontological foundations in that it enables it to take *full* account of human fallibility. We also argue that, in an economy, change and evolution can be viewed as generated *endogenously* by the interaction of two feedback mechanisms: (i) a negative feedback mechanism whereby market participants revise their previous expectations in the light of observed systematic discrepancies between expected and realized outcomes, and (ii) a feedback mechanism of ambiguous sign whereby market participants´ decisions and actions may alter the surrounding environment in an unpredictable way. The endogeneity of change in a market economy generates two main predictions: (i) macroeconomic equilibrium is fragile and *short-lived* thereby making it difficult for theorists and policy-makers to exploit it, and (ii) market economies exhibit boom-bust cycles of varying amplitude and duration so that there is *never* a repetition of the same cycle.

I. SOME REFLECTIONS ON POPPER´S APPROACH TO RATIONALITY AND ITS IMPLICATIONS FOR THE SOCIAL SCIENCES

`The physicist who is only a physicist can still be a first-class physicist and a most valuable member of society. But nobody can be a great economist who is only an economist — and I am even tempted to add that the economist who is only an economist is likely to become a nuisance if not a positive danger´ (F. A. Hayek, `The Dilemma of Specialization´, p. 123, in *Studies in Philosophy, Politics, & Economics*, 1967).

1. Introduction

Discussions by both philosophers of science and social science methodologists on Popper´s methodological proposal for the social sciences known as `Situational Analysis´ (*SA*) have focused either on its (in)compatibility with falsification (Hands, 1985, 1991, 1992; Notturno, 1998; Hedström *et al*., 1998) or on its interpretation.[1] In the former case, the debate has revolved around Popper´s confession (Popper, 1994, ch. 8) that his `rationality principle´ (*RP*), i.e., the notion that in the construction of models in the social sciences we must assume that actors´ behaviour is adequate or appropriate to their problem-situation (P-S), is *false* but nevertheless a *good enough approximation to the truth* (Popper, 1985).[2] This surprising confession by a philosopher of science whose academic reputation grew out of the formulation of a demarcation criterion for scientific hypotheses based on the requirement that the latter be potentially falsified has led some commentators to argue that Popper´s methodological proposal for the social sciences and, specifically, his *RP* is incompatible with the criterion of refutability as prescribed for the

natural sciences.[3] Some critics even argue that such incompatibility severely undermines Popper´s claim to methodological monism in the natural and the social sciences. Other critics have accused him of reintroducing a pure instrumentalism *à la* Friedman (1953) due to his methodological advice to immunize *RP* from potential refutation and of `contradicting his own explicit rejection of instrumentalist pretences to knowledge´ (Oakley, 1999, p. 32).

As for the interpretation of *RP*, the debate has centred on its role and status. For instance, Latsis (1983, p. 132) argues that `Popper´s account of the role and status of the rationality principle is obscure and unsatisfactory´. He shows that, in different parts of his work, Popper notes that *RP* is `almost empty´, `not a priori valid´, `clearly false´, `a good approximation to the truth´, and `the consequence of a methodological postulate´ (*op. cit.*, p. 133). According to him, the role of *RP* is to function as a `plastic interface´ between mental states and behaviour and this is the reason why it is declared by Popper to be false but close to the truth (*op. cit.*, p. 140). Specifically, *RP* is false if interpreted in a literal way because it does not *determine* behaviour in a `cast-iron´ fashion but is close to the truth because it captures the tendency of human behaviour to follow the logic dictated by P-S.[4] Crucially, he distinguishes between an `objectivist´ (*RPo*) and a `subjectivist´ version (*RPs*), and associates the former with Pareto (1917, section 150) and Parsons (1937, p. 58), and the latter with Popper himself. If *RPo* is adopted, the theoretician reconstructs the `objective´ P-S in an oversimplified way whereas, if *RPs* is adopted, P-S is reconstructed *as it is seen by the actors*.

Building on Latsis´ distinction between *RPo* and *RPs*, Nadeau (1993) discusses the role of *RP* in Popper (1985) and maintains that *RPo* is clearly false because actors´ behaviour is not always adequate to the `objective´ P-S whereas *RPs* is irrefutable and, hence, it can only be interpreted as a metaphysical statement (*op. cit.*, p. 459). He then

states that *RPs* is the correct interpretation since `the rationality principle that Popper puts at the theoretical core of all social sciences looks *more* like a "synthetic a priori truth or pure reason" in the domain of social reality than like an empirical law of nature´. Nevertheless, when asked to clarify whether *RP* is a `methodological principle´ or an `empirical conjecture´, Popper explains that `[t]his second case is precisely the one that corresponds to my own view of the status of the rationality principle: I regard the principle of adequacy of action (that is the rationality principle) as an integral part of every, or nearly every, testable social theory´ (Popper, 1994, p. 177). In other words, he views *RP* as an integral part of any empirical theory in the social sciences and, more specifically, as the animating part, just as the laws of motion of planets are an integral part of Newton´s gravitational theory. In an attempt to make sense of all this, Lagueux (2006, p. 203) argues that a methodological principle cannot be a part of a scientific theory whose constituent parts must be *empirical* rather than *a priori*. Yet, according to him, if *RP* cannot be a methodological principle, `the *decision* to immunize it can nonetheless be considered as based on a methodological principle´ (see Notturno, 1998, pp. 405-408). Although he uses a different terminology — methodological `rule´ instead of methodological `principle´ — de Bruin (2006, p. 213) also explains that the decision to adopt *RP* is a methodological rule according to which `one should always try to explain human behaviour in terms of reasons´. However, he adds that `there are good reasons to doubt whether the kind of principle of rationality that Popper discusses is really empirical at all… one could as well phrase it as a *metaphysical principle* that all actions have reasons´ (de Bruin, 2006, p. 216, emphasis added).

Despite the fact that some commentators have noted that there are two different notions of rationality in Popper´s philosophy, there has been very little discussion about the relation of *RP* and Popper´s evolutionary theory of knowledge and learning (PTKL).

The essence of Popper´s theory of knowledge is that all knowledge is conjectural and that we can *never* prove that a hypothesis is true albeit sometimes it is possible to prove that it is wrong. Likewise, the essence of Popper´s evolutionary theory of learning is that all living organisms (including human beings) learn by virtue of an imperfect and unending process that consists of subjecting their conjectures or hypotheses to trial and eliminating those ones which turn out to be erroneous while keeping *provisionally* those ones that are not falsified (Popper, 1963, p. 312).[5] In other words, our knowledge grows in a `negative´ sense by discarding erroneous conjectures through a process of trial and error-elimination. As a result of it, an implication of PTKL is that the most important feature of knowledge is its fallibility. A second element of Popper´s theory of learning is that the learning process is always *imperfect* insofar as it never reaches an optimum adaptation to the surrounding environment.

Now, a number of critics have referred to the apparent *duality* in Popper´s notion of rationality. To be sure, Popper (1985, p. 365; 1994, p. 181) distinguishes between rationality as a personal attitude — which he defines as the attitude of readiness to correct one´s beliefs when they turn out to be wrong — and his *RP* which, according to him, *has nothing to do* with the assumption that men are rational in this sense. Further, when he presents *SA* as his methodological proposal for the social sciences he writes that `when we speak of "rational behaviour" or of "irrational behaviour" then we mean behaviour which is, or is not, in accordance with the logic of the situation´ (Popper, (1966[1943a]), p. 97; 1944-45, sections 31 & 32). Be that as it may, Kerstenetzky (2009, p. 202) argues that the demarcation line between rationality and irrationality in Popper is the *incorrigibility* of one´s beliefs. That is, human behaviour is `rational´ if we are willing to correct our wrong beliefs and `irrational´ if otherwise. Similarly, Lagueux (2006, p. 202) points out that it is the `tendency to correct oneself by criticism´ that represents true

rationality in Popper whereas Vanberg (2002, p. 19) remarks that, on the one hand, Popper presents *RP* as the methodological foundation for the social sciences but, in other parts of his work, he sketches out a framework for the analysis of human behaviour that relies on a different approach at purposeful behaviour that he defines as `conjecture-based problem-solving behaviour´. Finally, the conflict between the view of human agency depicted in PTKL and *SA* is also noted by Oakley (1999, p. 25; 2002, p. 468).

The main purpose of this essay is to analyse the compatibility or otherwise of these two seemingly different notions of rationality in Popper´s work. In the process, we will make five claims. Our first claim is that there is a certain tension between PTKL and *SA* when their relation is analysed from the standpoint of the `rationality of the agents´ whose behaviour the theoretician seeks to capture in a situational model albeit the tension disappears when the relation is treated from the standpoint of the `rationality of the theoretician´. Our second claim is that the nature of the tension between PTKL and *SA* depends on whether the theoretician adopts the `objectivist´ or the `subjectivist´ version of *SA*. In particular, we will argue below that the tension between PTKL and the `subjectivist´ version stems from the fact that, in the latter, it is implicitly assumed that agents´ view of P-S is, at least partially, wrong which implies that agents do not tend to eliminate their mistakes as PTKL posits. By contrast, we will argue that the tension between PTKL and the `objectivist´ version of *SA* stems from the fact that: (i) if agents behave according to PTKL it is not necessarily the case that their decisions will be adequate or appropriate to the `logic of the situation´ because the former *only* implies that agents tend to eliminate their (past) mistakes and, hence, in the wake of changes in the surrounding environment agents´ decisions may be inadequate to the `logic of the (new) situation´, and (ii) adoption of the `objectivist´ *SA* implies *de facto* the imposition of the theoretician´s view of P-S on agents´ but it is unlikely that if agents behave according to

PTKL their view of P-S will converge to the theoreticians´. Our third claim builds on the ideas of Hayek (1943) about the nature of the `facts´ of the social sciences and is that, in the way it is presented by Popper and his commentators, the `objectivist´ version of *SA* represents a *limit* case which presupposes that P-S is (fully) *independent* of agents´ beliefs. Our fourth claim is closely related to the previous one and consists of the idea that, if Hayek´s ideas are accepted, it follows that the natural strategy for social scientists is to seek to reconstruct P-S *as agents´ see it*. Our fifth and last claim is that, unlike what Popper and some of his commentators suggest, the difference between the `objectivist´ and the `subjectivist´ version of *SA* is not that in the former the theoretician reconstructs P-S *as it actually is* whereas in the latter she does it as *agents see it* but, rather that in the former the theoretician seeks to reconstruct P-S *as she sees it* whereas in the latter she does it *as she believes agents see it*. The content of the essay is as follows. The following section introduces PTKL. In section 3, we expound Popper´s *SA*. In section 4, we discuss the duality in Popper´s notion of rationality by addressing: (i) the implications of adopting either the `objectivist´ or the `subjectivist´ version of *SA*, (ii) Hayek´s ideas about the `facts´ of the social sciences, (iii) a reformulation of the `objectivist´ and `subjectivist´ version of *SA* that takes on board Hayek´s ideas on the methodology of the social sciences, and (iv) the division line between `rational´ and `irrational´ behaviour in PTKL and *SA*. Finally, section 5 summarizes and concludes.

2. Popper's evolutionary theory of knowledge and learning

Inductive inference is reasoning from the past observed behaviour of objects to their future behaviour. The 'problem of induction' was originally raised by David Hume (2006[1748]) who pondered whether inductive evidence can go beyond the available evidence in order to predict future events. He argued that past evidence can tell us only

about past experience. Hume's main argument was that we cannot rationally justify the claim that nature will continue to be uniform merely because it has been in the past, as this is using the sort of reasoning (i.e., induction) that is under question, that is, it would be *circular reasoning*. Hume (*op. cit.*) also noticed that we *tend* to believe that phenomena behave in a regular fashion, that is, that certain patterns in the behaviour of objects persist into the future.

Now, Popper defines the philosophical 'problem of induction' as the problem of providing a rational justification for the common belief that the future will be (largely) like the past (Popper, 1972, p. 2). He identifies two problems in Hume´s criticism of induction: (i) a logical problem (*HL*), and (ii) a psychological problem (*HP*). First, Hume´s *HL* is whether we are justified in reasoning from repeated instances of which we have some experience to other instances of which we have no experience. Hume´s answer to *HL* is negative no matter how many repetitions of the instances there are. Second, Hume´s *HP* is why, notwithstanding it, reasonable people believe that instances of which they have no experience at all will tend to conform to those of which they have experience. Hume´s answer to *HP* is that 'the psychological mechanism of association forces them to believe, by custom or habit, that what happened in the past will happen in the future' (*op. cit*., p. 90). According to Popper, this explains why Hume abandoned rationalism and viewed man as a product of blind habit. Specifically, Hume (1875, pp. 41-46) argues that *custom* is the means by which nature induces reasonable people to accept the existence of `constant conjunctions´ of events.[6] By contrast, Popper argues that there is no such thing as induction by repetition either in psychology or science:

'We do not act upon repetition or "habit", but upon the best tested of our theories which… are the ones for which we have good rational reasons; not of course good reasons

for believing them to be true, but for believing them to be the best available from the point of view of a search for truth or verisimilitude… The central question for Hume was: do we act according to reason or not? And my answer is: Yes.' (*op. cit*., p. 95)

Popper restates Hume´s problem of induction as follows. First, he denies that a theory can be simply justified by assuming the truth of certain observation statements. Rather, he insists that all theories are hypotheses and, hence, they can be overthrown (*op. cit*., p. 13). Further, he states that paradoxically induction is *inductively invalid*, that is, the claim that induction is a legitimate way to acquire (true) knowledge needs to be supported by a 'higher' principle that has, in its turn, been established inductively. But this strategy ultimately leads us into an *infinite regress* insofar as we will endlessly need to resort to a superior principle that has been discovered through induction. Second, he puts forward the proposition that the claim that an explanatory universal theory is false can be justified by the truth of certain observation statements (*op. cit*., p. 7). As the typical example goes, no matter how many white swans we come across, the finding of just one black swan will lead to the rejection of the universal statement 'all swans are white'. Consequently, he urges scientists to construct *severe* tests that help detect false theories so that, by a method of *elimination*, they may eventually hit upon a true theory even though we can never establish its truth (*op. cit*., pp. 14-15).[7] Thus, he argues that there is an *asymmetry* between verification and falsification; any conjecture may be true or false but even if it turns out to be true, there is no way we can *ever* prove it (*op. cit*., p. 12). According to Popper, the method of science is 'the method of bold conjectures and ingenious and severe attempts to refute them' (*op. cit*., p. 81). Since all theories involve universal statements, we can only 'learn' by proving that our knowledge is false. Specifically, learning takes place either when we reject one´s prior theory or when we are

forced to adjust one´s theory in a way that recognizes that in its prior version it was false (*op. cit*, p. 81). In short, Popper´s ideas on scientific methodology can be seen as a sub-product of PTKL:

`Although I shall confine my discussion to the growth of knowledge in science, my remarks are applicable without much change, I believe, to the growth of pre-scientific knowledge also — that is to say, to the general way in which men, and even animals, acquire new factual knowledge about the world. The method of learning by trial and error — of learning from our mistakes — seems to be fundamentally the same whether it is practised by lower or by higher animals, by chimpanzees or by men of science. My interest is not merely in the theory of scientific knowledge, but rather in the theory of knowledge in general. Yet the study of the growth of scientific knowledge is, I believe, the most fruitful way of studying the growth of knowledge in general. For the growth of scientific knowledge may be said to be the growth of ordinary human knowledge *writ large´* (Popper, 1963, p. 216).

and, elsewhere, he writes:

`Organisms evolve by trial and error, and their erroneous trials — their erroneous mutations — are eliminated, as a rule, by the elimination of the organism which is the "carrier" of the error. It is this part of my epistemology that, in man, through the evolution of a descriptive and argumentative language, all this has changed radically. Man has achieved the possibility of being *critical of his own tentative trials, of his own theories*. These theories are no longer incorporated in his organism, or in his genetic system: they

may be formulated in books, or in journals; and they can be critically discussed, and shown to be erroneous, without killing any authors or burning any books: without destroying the "carriers"… *critical reason is the only alternative to violence so far discovered´* (Popper, (1966[1943b]), p. 292).

Third, Popper argues that theories are genetically incorporated into all our sense organs and this predisposes us to discriminate *a priori* between relevant or absorbable input and input that can be ignored (Popper, 1972, p.72). For instance, sense organs like the eye *only* react to those selected environmental events which they 'expect'. However, according to him, this prior knowledge cannot be the result of observation; it must be the result of adaptation to the surrounding environment by trial and error. He claims that 99 percent of the knowledge of all living organisms is inborn and incorporated into our biochemical constitution (Popper, 1990, p. 46). Furthermore, he argues that there is no theory-free language to help us interpret external data because primitive theories emerge together with language. Therefore, there is no such thing as pure perception since all languages are *theory-impregnated* (Popper, 1972, p. 145). This leads him to reject any epistemology which chooses our 'direct' observations and perceptions as the starting point; the fact that theories are built into our sense organs implies that 'the epistemology of induction breaks down even before taking its first step' (*op. cit*, p. 146).

Lastly, Popper´s rejection of Hume´s *inductive* theory of beliefs formation leads him to maintain that 'logical' considerations may be duly transferred to 'psychological' considerations. According to him, not only do we reason rationally and thus contrary to the principle of induction, but *we also behave rationally*. He labels this the 'principle of transference' (*op. cit*, p. 6). By applying this conjecture to human psychology he then arrives at the method of *trial and error-elimination* in which the trials correspond to the

formation of competing hypotheses whereas the elimination of errors corresponds to the refutation of (false) hypotheses. In other words, he proposes the theory that *individuals do not really think in an inductive way but rather form their beliefs by eliminating false hypotheses*. The theory of knowledge and learning that thus emerges is *evolutionary*. However, such theory implies that adaptation is always 'imperfect':

'Some of the errors that have entered the inheritable constitution of an organism are eliminated by eliminating their bearer; that is, the individual organism. But some errors escape, and this is one reason why we are all fallible: our adaptation to the environment is never optimal, and it is always imperfect' (Popper, 1990, p. 47).

Further, Popper asserts that *no equilibrium state of adaptation* is reached by the application of the method of trial and error-elimination since (i) no optimal trial solution to any specific problem is likely to be offered, and (ii) the emergence of new structures and instructions involves a continuous *change* in the environmental situation (Popper, 1994, p. 4). More specifically, and crucially, Popper presents the growth of knowledge as bringing in its wake changes in the surrounding environment:

'Our very understanding of the world changes the conditions of the changing world; and so do our wishes, our preferences, our motivations, our hopes, our dreams, our fantasies, our hypotheses, our theories. Even our erroneous theories change the world, although our correct theories may, as a rule, have a more lasting influence. All this amounts to the fact that *determinism is simply mistaken*' (Popper, 1990, p. 17).

In short, Popper makes it clear that the past *affects* but does not determine the future, i.e., the future is not pre-determined. That is, the future is 'objectively open' (*op. cit*., pp. 17-18). As noted in Vanberg (2002, p. 8), Popper´s theory of knowledge and learning exhibits a remarkable similarity to the arguments developed by biologist Mayr (1988) in the sense that both Popper and Mayr argue that intentional problem-solving behaviour can be interpreted as behaviour governed by programs or conjectures which are the product of evolutionary learning by trial and error-elimination. As Vanberg (*op. cit*.) explains, this approach implies that 'there is a continuum from the behaviour of primitive organisms, governed entirely by genetically coded programs, to the sophisticated, deliberated actions of "rational man" governed by conjectures or mental models that are stored in memory'. According to Vanberg (*op. cit*., p. 27), 'even the most deliberate and conscious instances of problem-solving are no less "program-based" than any subconscious or unconscious routine behaviour, in the sense that they, too, have nothing else to rely on than *conjectures*…´. Thus, Vanberg views rationality as a problem-solving capacity that is stored on a person's catalogue of conjectures or programs that exhibits no more wisdom than that embedded in the knowledge acquired in the past. According to this view, rationality 'cannot guarantee pre-adaptedness, it is instead a matter of the backward-looking adaptedness of behavioral programs that allows for a tentative, forward-looking response to current problem-situations' (*op. cit*., pp. 16-17). As we will see, this aspect of the notion of human rationality embedded in both PTKL and Mayr´s notion of problem-solving behaviour implies the ability to solve certain problems that agents have encountered in the past does not necessarily imply that they are endowed with the ability to solve new (and different) problems that they encounter as the surrounding environment changes.

Let us address Popper´s distinction between *subjective* and *objective* knowledge. The former consists of certain inborn dispositions of organisms and of their acquired modifications to act, whereas the latter consists of the logical content of our theories and, as such, it includes the world of language, conjectures, arguments, and scientific theories.[8] As for subjective knowledge, Popper´s diagnosis is that it is part of a complex but accurate apparatus of adjustment that, in the main, works like objective conjectural knowledge, namely, by the method of trial and error-elimination or 'auto-correction' (Popper, 1972, p. 77). As for objective knowledge he notes that only a tiny part of it can be given sufficient reasons for certain truth. Such tiny part is denoted as *demonstrable* knowledge and comprises the propositions of formal logic, and arithmetic. All else, including knowledge associated to the natural and the social sciences, is conjectural or hypothetical knowledge and, hence, there are no sufficient reasons for holding it to be true (*op. cit*., p. 75). Thus, from the point of view of objective knowledge, all theories are *conjectural* albeit this does not preclude the possibility that some of them are true.

The method of science, according to Popper, is 'the method of bold conjectures and ingenious and severe attempts to refute them' (*op. cit*., p. 81). Since all theories involve universal statements, we can `learn´ by proving that our knowledge is false. Specifically, learning takes place either when we reject one´s prior theory or when we are forced to adjust one´s theory in a way that recognizes that in its prior version it was false (Popper, 1972, p. 81). Thus, we can only 'learn' by refuting our prior knowledge claim. As noted in Boland (2003*b*, p. 242), an implication of PTKL is that the mere accumulation of information does not increase the odds that our theories happen to be true because, as Popper insists, all we can 'learn' from experience is that some of our theories are false. In this respect, Boland (2003*b*, p. 248) makes a useful distinction between the quantitative and qualitative views of knowledge. The former corresponds to the so-called 'bucket

theory of knowledge' whereas the latter corresponds to Popper´s theory of knowledge. He then proposes the metaphor that, in Popper´s Socratic theory of learning, 'knowledge is more like health that one can improve than wealth that one can have more of'. Hence, according to Boland (*op. cit*.), learning consists of improving one´s knowledge rather than of increasing it.

The distinction between objective and subjective knowledge also leads Popper to distinguish between three different worlds or ontological domains: (i) the world of physical objects or states (World 1), (ii) the world of states of consciousness, or of mental states (World 2) and, lastly, (iii) the *autonomous* Platonic-like world of objective contents of thought, especially of scientific thoughts and works of art (World 3).[9] His main thesis on this respect is that almost all our subjective knowledge (belonging to World 2) depends upon World 3, that is, on *linguistically formulated* theories (Popper, 1972, p. 74). However, he argues that our mind may be connected to objects of both World 1 and 3 and this allows World 2 to act as a *mediator* between them. Further, he notes that World 3 exerts a profound influence upon World 1 through the actions of technologists who implement changes in World 1 by applying the predictions of these theories. Finally, he argues that we always select our P-S against a World 3 background which consists of, at least, a language and that 'the activity of understanding consists essentially in operating with third-world objects' (*op. cit*., p. 164). In particular, the development of science and art presupposes the prior existence of the human language which leads Popper (1990) to argue that the latter is, by far, the most important product of the human mind:

'Language makes it possible to consider our theories critically: to look at them as if they were external objects, as if they belonged to the world outside of ourselves which

we share with others. Theories become objects of criticism, like the beaver dam' (Popper, 1990, p. 51).

Next, Popper sees science as one of the greatest creations of the human mind, comparable only to the emergence of a descriptive and argumentative language, since its creation allowed men to replace: (i) the elimination of error in the violent struggle for life by non-violent rational criticism, and (ii) killing (World 1) and intimidation (World 2) by the impersonal arguments of World 3 (*op. cit*., p. 84). He defines epistemology as the theory of the growth of *scientific knowledge*, that is, the theory of problem-solving, or of the critical discussion, evaluation, and critical testing of competing theories (*op. cit*., p. 142). However, as we mentioned above, PTKL is not only applicable to scientific knowledge but to any type of knowledge. As such, he sees scientists acting on the basis of hunches and guesses concerning what looks promising for future growth in the third world of objective knowledge. In so doing, he identifies content and virtual explanatory power as the most important criteria for the *a priori* appraisal of theories where both are related to their degree of testability. In turn, the most important criterion for their *a posteriori* appraisal is 'verisimilitude' or 'nearness to truth' and this, he argues, depends upon the way a theory has stood up to severe tests (*op. cit*., p. 143).[10] The evaluation process is always critical and aims at error-elimination. Lastly, the exposure of scientific hypotheses to severe tests and criticism from the scientific community guarantees their increasing accuracy at explaining phenomena:

`What is characteristic of science is that the selective system which weeds out among the variety of conjectures involves deliberate contact with the environment

through experiment and quantified prediction, designed so that outcomes quite independent of the preferences of the investigator are possible. It is pre-eminently this feature that gives science its greater objectivity and its claim to a cumulative increase in the accuracy with which it describes the world´ (Campbell, 1974, p. 434).

Finally, the growth of World 3 is not a repetitive or cumulative process alike Lamarckian instruction but a Darwinian selection process which consists of systematic error-elimination (*op. cit.*, p. 149; also Popper, 1994, ch.1). He identifies three different levels of adaptation: genetic, behavioural learning, and scientific discovery. Scientific discovery is, according to him, a special case of adaptive behavioural learning. Popper asserts that, on all levels, the mechanism of adaptation to the surrounding environment is essentially the same, i.e., a Darwinian selection process by trial and error-elimination. In short, he views science 'as a means used by the human species to adapt itself to the surrounding environment: to invade new environmental niches, and even to invent new environmental niches' (Popper, 1994, p. 2).

3. Popper's methodological prescription for the social sciences

We now address Popper´s methodological prescription for the social sciences known as *SA* and the status of *RP*. Early presentations of the method of *SA* can be found in Popper´s *Open Society* (Popper, (1966[1943a], ch. 14, especially p. 97), in his *Poverty of Historicism,* originally published in three articles in *Economica* and, then, as a book (Popper, 1944-45, sections 31 & 32), in a French paper (Popper, 1967), and in *Objective Knowledge* (Popper, 1972, p. 179). However, the place where he presents it thoroughly is in the article titled "Models, Instruments, and Truth: The Status of the Rationality

Principle in the Social Sciences" (Popper, 1994, ch. 8). This book chapter was originally written in response to an invitation that Popper received in the early 1960s from the Department of Economics at Harvard University and the lecture he delivered there on 26 February 1963. As noted in de Bruin (2006, footnote 1), in 1963 and 1964 two new sections were added and a small extract was then circulated in the London School of Economics in 1967 and 1968. This extract was translated into French and published as `La rationalitè et le statut du principe de rationalitè´ (Popper, 1967) and, then, a Spanish translation of the French translation appeared about a year later. A revised version of the English extract was published in 1983 on pages 357-365 of an anthology titled *A Pocket Popper* which is currently available in *Popper Selections* (Popper, 1985). However, the full text of the speech at Harvard University was not made available until 1994 when it was published in a collection of Popper´s essays titled *The Myth of the Framework.*

3.1. The Rationality Principle

Popper´s thesis in that chapter is that there is no fundamental difference between the natural sciences and the social sciences since both of them resort to the construction of models or *typical* P-S to explain and predict events. If anything, models are viewed by him as being even more important in the social sciences due to the non-existence of universal laws. In any case, he argues that the models of the theoretical social sciences are always an *over-simplification* of reality and, hence, do not represent the facts truly. According to him, the fundamental problem of the social sciences is 'to explain and understand events in terms of human actions and social situations' (Popper, 1994, p. 166). In turn, the reconstruction of social situations should include the consideration of the relevant 'social institutions' which he defines as 'all those things which set limits or create obstacles to our movements and actions' (*op. cit*., p. 167). In his autobiography, Popper

makes it clear that his methodological proposal for the social sciences stems from an 'attempt to generalize the method of economic theory (marginal utility theory) so as to become applicable to the other theoretical social sciences' (Popper (2002[1976a], p. 135).

Next, Popper makes a distinction between 'rationality' as a personal attitude and his *RP*. In particular, he makes it clear that his *RP* has nothing to do with the assumption that men adopt a rational attitude. Rather, he defines it as an *a priori* methodological principle which assumes that *our actions are adequate to our problem-situations as we see them* (Popper, 1994, p. 181). More specifically, he remarks that *RP* is *not* true: 'The rationality principle is false. I think there is no way out of this. Consequently, I must deny that it is *a priori* valid' (Popper, 1985, p. 361).[11] Notwithstanding it, he believes it represents a good approximation to the truth. Thus, *RP* 'does not play the role of an empirical explanatory theory, of a testable hypothesis' (*op. cit*., p. 360). Rather, he views it as an integral part of every testable theory and proposes to avoid blaming it whenever our theory breaks down in the wake of empirical tests. His methodological advice to social scientists is thus never to abandon *RP* so that, in the wake of a refutation of their model, they should always revise their models of the agent's P-S.[12]

As Koertge (1975) shows, Popper's views on the *RP* have evolved over time. As time passed by, he tended to weaken his claims about the kinds of actions that agents could be expected to perform so that 'where he had earlier spoken of actions as being `rational´ or `appropriate´, he now characterized them as `adequate´, or `adapted´, or `in accordance with the situation´ (*op. cit*. p. 441). According to him, the most likely reason for this evolution in terminology was his increasing emphasis on the fact that the P-S which played a central role in the explanation was not so much the agent's *objective* P-S but, rather, the agent's *theory* of her P-S or the P-S *as the agent saw it* (Koertge, 1975, p. 442). She explains that *RP* really consists of two clauses: the first (*RP-1*) says that 'every

action (by a person) is a rational response to some problem-situation' whereas the second (*RP-2*) tells us that 'every person in a problem-situation responds rationally to it' (*op. cit*., p. 443). In turn, *RP-1* entails: (i) that the response was issued through a *methodical* appraisal of the set of possible solutions available to the actor, (ii) that a description of both P-S and the appraisal process could be verbalized by the actor, and (iii) that the person acted as she did as a result of the appraisal process so that if a better alternative had been available to her she would have taken it. Thus, the complete *RP* formulated in Koertge (1975) emphasizes the close connection between the action and the *systematic deliberation* process from initial conditions that made the agent behave as she did. Further, Koertge (1979, p. 90) points out that requirement (i) above implies that 'some systematic non-random decision procedure be used' albeit she notices that Popper did not specify the minimal requirements which acceptable decision rules should satisfy. This means that *RP* can, in principle, be supplemented with different theories of belief formation. As Koertge (*op. cit*., p. 92) explains, for Popper, to explain an action using *RP* does not 'imply that the agent's beliefs are reasonable nor even that her way of making decisions is the best possible one' but only presupposes that agents assess the situation in a *systematic* way.

Next, as we noted above, Popper´s methodological advice to social scientists is never to abandon *RP* so that, in the wake of a refutation of their model, they should always revise instead their model of the agent's P-S. He offers two arguments in favour of this strategy: (i) that we learn more if we blame our situational model, and (ii) that the adoption of *RP* 'reduces considerably the arbitrariness of our models' (*op. cit*., p. 362). As for the first argument, he explains that:

‘The main argument in favour of this policy is that our model is far more interesting and informative, and far better testable, than the principle of the adequacy of our actions. We do not learn much in learning that this is not strictly true: we know this already’ (Popper, 1985, p. 362).

Likewise, Caldwell (1991, p. 25) argues that, although immunizing stratagems should be generally avoided ‘at least in the special case of situational analyses, one is able to *criticize more severely and obtain fruitful criticisms* if one blames the model rather than *RP* whenever a falsification occurs’. As for the second argument, Popper explains that:

‘The attempt to replace the rationality principle by another one seems to lead to complete arbitrariness in our model-building. And we must not forget that we can test a theory only as a whole, and that the test consists in finding the better of two competing theories which may have much in common; and most of them have the rationality principle in common’ (Popper, 1985, p. 362).

As Hands (1985, p. 87) remarks, Popper’s first argument above means that if we are consistent with *RP* ‘the falsification of a specific theory only means that we have misspecified the “situation”, i.e., that we have attributed the wrong preferences or constraints to the individual’. In turn, Popper’s second argument implies that although *RP* is potentially falsifiable *we choose to make a methodological decision that, when faced with a falsifying observation, we will stick to it* and revise instead our hypotheses about the desires, beliefs, and constraints faced by agents (Hands, 1985, p. 88). Notably, Becker

(1976) resorts to a similar argument to justify the use of rational choice theory. According to him, human behaviour can be viewed from the standpoint of individuals who seek to maximize their utility from a stable set of preference and subject to a given constraint. Where action appears to deviate from the predictions of neoclassical utility theory, Becker claims that little is gained from resorting to explanations in terms of irrationality, changes in preferences or cultural values, etc… for such explanations are *ad hoc* and may even be contradictory. Furthermore, he adds that the question is left unanswered of just *why* human behaviour should be sometimes rational but sometimes not.

According to Caldwell (1991, p. 15), there are two main weaknesses in Popper´s presentation of *SA*: (i) vagueness about how it should be implemented, and (ii) Popper´s apparent belief that *SA* is the only adequate method to adopt in the theoretical social sciences. As for the first point, we have presented above a clearer explanation of how to apply it suggested in Koertge (1979). As for the second point, Caldwell (1991, p. 16) readily admits that *SA* is a powerful and fruitful method for the social sciences, yet he criticizes Popper´s idea that *SA* is the *only* legitimate method for the theoretical social sciences. Further, he recognizes that there is a tension between falsificationism and *SA* owing to the fact that *RP* adopts the status of a methodological prescription that plays the role of an immunizing stratagem.[13] A discussion of this issue is in Hands (1985, p. 89) who argues that, by Popperian standards, scientific explanations based on *RP* 'are as close to metaphysical explanations as they are to scientific explanations' and, hence, the tension between these two methodological principles can hardly be resolved. Caldwell (1985) proposes to solve the conflict between *SA* and falsificationism by adopting a broader conception of acceptable scientific practice based on 'critical rationalism' whose goal is to subject theories to an *optimal* amount of criticism. In turn, the latter will depend on both the specific problem to be solved and the nature of the problem under investigation

(*op. cit*., p. 25). Such prescription was proposed in Klappholz & Agassi (1959) and, later on, it has been promoted by Boland (2003*a*) who stresses that the only generally applicable methodological rule is the exhortation to be always critical and ready to subject one's hypotheses to critical scrutiny. More specifically, he insists we should focus on the Socratic-Popper identified in Klappholz & Agassi (1959) and thus discard the Lakatos-Popper (also known as Popper the 'falsificationist') promoted by Latsis (1972) and Blaug (1975). According to him, if we put falsificationism aside in favour of 'critical rationalism' the conflict between *SA* and falsificationism vanishes.

3.2. The two versions of the Rationality Principle

Latsis (1983) was probably the first commentator to identify the existence of an `objectivist´ version (*RPo*) and a `subjectivist´ version (*RPs*) of *RP* in Popper´s work.[14] In the former, the relevant P-S is that one as seen by the theoretician whereas, in the latter, the theoretician is supposed to reconstruct P-S as seen by agents. Latsis (*op. cit*.) denotes the former as the `strong´ version of *RP*. According to Latsis (*op. cit*., p. 131), Popper both *weakens* and *widens* the notion of rationality in human behaviour when adopting *RPs*.[15] Building on the distinction between *RPo* and *RPs*, Nadeau (1993, p. 463) notes that `an attentive reading of the 1967 text shows that although Popper views the *RP* as an explanatory principle throughout the text, he surreptitiously changes his way of formulating it during the course of his argument, going from an objectivist formulation at the beginning of the text to a subjectivist formulation at the end´. Hands (1991, footnote 14) recognizes that `Popper is really unclear on this´, and Latsis (1983, p. 133) claims that Popper seems either `confused or deliberately elusive´ on this issue. Be that as it may, Hands (*op. cit*.) points out that in his 1985 text Popper adopts the subjectivist interpretation when he openly says that rationality is only `as agents see it´ and *SA* can

thus be applied to apparently irrational behaviour such as the behaviour of a `madman´ (Popper, 1985, p. 363).[16] However, he adds that Popper also denotes *SA* `a purely objective method´ which `can be developed independently of all subjective and psychological ideas (Popper, 2002[1976*a*], p. 172) and that, elsewhere, Popper says that *RP* is the `general law that *sane* persons as a rule act more or less rationally´ (Popper, 1966, p. 265).

Now, in a passage of his 1967 French paper, Popper (1985, p. 363) proposes his famous example of the `flustered driver´ who, by trying to park stubbornly his car in evidently too small a space, does not act in a way that is appropriate to the situation *in which he finds himself* and then recognizes that `we employ the rationality principle to the limit of what is possible whenever we try to understand the action of a madman´ (Popper, 1994, p. 179). It is in the section of the chapter where he notices that cases of neurosis have been explained by Freud and other psychologists with the help of their own version of the *RP* that he switches to a *subjectivist* version of *RP*. Then, in a key note to one of the paragraphs (footnote 19), he acknowledges that he refers successively to two versions of his *RP* and even identifies a third intermediate version according to which P-S is said to be `as the agent could (within the objective situation) have seen it´ (Popper, 1994, ch. 8, footnote 19).[17] In the aftermath of it, Lagueux (2006, p. 201) concludes that, according to Popper, `what the agent sees may or may not be considered a part of the objective situation that the model describes´. Summing up, the `objectivist´ version (*RPo*) supposes that agents possess `true´ knowledge; the `subjectivist´ version (*RPs*) supposes that the alleged knowledge that agents possess is *partially* wrong; and the third version constitutes an intermediate case. However, in all three versions of *RP* it is assumed that the agent acts in a way that is appropriate to the state of *his* knowledge (Popper, 1994, ch. 8, footnote 19; Lagueux, 2006, p. 201).

Next, building on the terminology coined in Latsis (1972), Kerstenetzky (2009, p. 201) denotes *RPo* the `maximal´ or `single-exit´ interpretation and *RPs* the `minimal´ or `multiple-exit´ interpretation of *RP*. The `single-exit´ interpretation stems from the fact that, if it is supposed that the agent perceives P-S in an objective way, there is thus only `one´ possible solution whereas the `multiple-exit´ interpretation captures the idea that, in principle, there are as many solutions as subjective perceptions of P-S exist. It is the `multiple-exit´ interpretation that is of interest in the context of the `subjectivist´ *SA.* In particular, the issue is *whether* we can assume for methodological purposes that the different subjective perceptions of P-S held by actors actually converge on a `single´ one and, if so, *how* this convergence comes about. Alike Jacobs (1990), Kerstenetzky (*op. cit*) associates the `objectivist´ or `single-exit´ modelling to the influence on Popper of the work of Weber. By contrast, Hedström *et al*. (1998, p. 359) do not think there is textual evidence that Popper got the inspiration for the notion of *SA* from Weber´s work and suggest that if there was any influence at all it was probably *indirect* since Hayek — a friend of Popper — admired Weber. However, it could be argued that Popper´s method for the theoretical social sciences takes on board Weber´s notion of `interpretive understanding´ or `*verstehen´* — developed later on by the Austrian economists — and, especially, his notions of `ideal type´ and of `instrumental rationality´, i.e., the use of rationality to bring about change in the surrounding world in the interest of the actor (Weber, 1949).[18] Be that as it may, there is some textual evidence that points to Hayek as the most important *direct* source of influence on Popper´s work. Notably, Popper (1994, ch. 8, note 1) writes that `I was particularly impressed by Hayek´s formulation that economics is the "logic of choice"´ as expressed in his essay titled `Economics and Knowledge´ (Hayek, 1948[1937], pp. 33ff). According to Popper, it was this that led him to the formulation of the `logic of the situation´ in his *Poverty of Historicism*.[19]

Finally, there is the issue of the status of *RP*. We have already mentioned above the profound ambiguity of Popper´s explanation about the status of RP. The subsequent discussion about the role and status of *RP* among Popper´s commentators focused on the distinction between *RPo* and *RPs*. For instance, Lagueux (1993, 2006) argues that, even if we adopt the `subjectivist´ interpretation, *RP* cannot be *a priori* true because, according to him, it is simply not true that people always act appropriately according to the P-S *as they see it*. Notwithstanding it, he thinks that *RP* occupies an exceptional place in the social sciences because it constitutes a *condition of intelligibility* of any phenomenon that derives from human action. More specifically, the latter can only be intelligible, i.e., understood by an external observer, when it is motivated by reasons, that is, when it represents an appropriate response to P-S as seen by the agent (Lagueux, 2006, p. 205). He concludes that maintaining *RP* after acknowledging that it is not *a priori* true is, after all, to claim that `in spite of the fact that irrational decisions occur, human actions are nonetheless normally understandable´ (*op. cit.*).

4. The notion of rationality in Popper´s philosophy of the social sciences

In section 2 we showed that PTKL implies that: (i) all knowledge is conjectural, (ii) that we learn through an (endless) process whereby we subject our conjectures to trial and discard those ones that turn out to be wrong, and (iii) that the learning process is imperfect and never converges to an optimum. Consequently, the most important feature of knowledge is its fallibility. By contrast, Popper´s methodological proposal for the social sciences has been denoted as 'situational determinism' (Latsis, 1972; Oakley, 2002) which suggests that there may be some key epistemological differences between PTKL and *SA*. The first thing we should like to note is that PTKL is a theory about the nature of knowledge and its growth over time while *SA* is a methodological prescription

aimed, arguably, at speeding up the rate of progress of the social sciences so that these two elements of Popper´s philosophy do correspond to the positive and methodological domain respectively. That said, we believe there is also a normative element in PTKL since trial and error-elimination can also be said to be the way we *should* behave when seeking to expand our knowledge. In any case, error-elimination can only proceed *after* there is clear-cut evidence that, retrospectively, a decision made in the past was wrong. However, this does not provide us with a systematic rule for making decisions *in the future* other than to avoid repeating the same mistakes made in the past. In short, error-elimination is an incomplete guide to decision-making.

Next, we may wonder how *SA* would look like if the agents that are the object of the modelling exercise exhibited a theory of knowledge and learning akin to PTKL. To be sure, the situational model of the typical P-S consists of three elements: (i) external (and observable) elements such as the physical and social constraints agents are subject to, (ii) the knowledge and information that agents possess, and (iii) their goals and aims. Now, if agents behave according to PTKL, then the situational model of the typical P-S should incorporate the knowledge they possess which would include the experience accumulated from mistakes they made in the past given the specific circumstances that prevailed at that time. Therefore, adequate behaviour would imply, as a minimum, not repeating previous mistakes. However, as we noted above, there is no further guidance for agents stemming from PTKL as far as future decision-making is concerned in case they encounter new (and different) P-S. In short, PTKL appears to be compatible with *SA* provided the situational model includes agents´ learning from previous mistakes.

4.1. PTKL versus *SA*: the `rationality of agents´

We noted above that several commentators, as well as Popper (1994, ch. 8, note 19) himself, identify two different versions of *RP*: an `objectivist´ version (*RPo*) and a `subjectivist´ version (*RPs*). According to the former, the relevant P-S is the `objective´ P-S, that is, the P-S *as it actually is* whereas, according to the latter, the theoretician should reconstruct P-S *as it is actually seen by the agents*. As Popper (1972, p. 179) recognizes, in both cases P-S is *conjectured*.[20] That said, we will argue below that, if Hayek´s ideas on the nature of the `facts of the social sciences´ are taken on board, there is no reason *a priori* to expect that the theoretician´s view of P-S is *closer* to the `true´ P-S than agents´ (Hayek, 1943). This is because, as Hayek argues, the theoretician does not possess superior relevant knowledge that is not shared by agents. Be that as it may, *RPo* and *RPs* constitute two different modelling strategies in the social sciences the consequences of which, to the best of our knowledge, have not been explored so far. Notably, an antecedent is Schumpeter´s distinction between `objective rationality´ and `subjective rationality´ (Schumpeter, 1984). He defines the former as consisting of the `applicability of a rational schema to the actor´s behaviour´ and he defines the latter as the `conformity of the actors´ mental processes to a rational schema´ (*op. cit.*, p. 583). Crucially, he states that the former need not imply the latter and criticizes the tendency of some social scientists to *implicitly* identify the rationality of the `observer´ with the subjective rationality of the `observed´ (*op. cit*., p. 583). He uses the example of the neoclassical theory of monopoly to illustrate the notion of `objective rationality´:

`The model just described is the product of the analyst´s mind as much as any physical theory is, and does not in itself say anything about reality or about anybody´s actual behavior or rationality… Even if the model should fit anyone´s behaviour this does

not mean that the individual in question consciously aims at the result and still less that he arrives at it by processes at all similar to the analytic procedure´ (Schumpeter, *op. cit.*, p. 580).

Schumpeter´s notion of `objective rationality´ is closely associated to his notion of `rationality of the observer´ whereas the notion of `subjective rationality´ is coupled to his notion of the `rationality in the observed´. In the example of monopoly theory, he explains that the construction of a model will give us the conditions under which the maximization of profits will be attained thereby setting up a standard against which the theorist can compare actual behaviour. However, he makes it clear that such model is *entirely* a product of the `rationality of the observer´ and, hence, the usefulness of the modelling exercise will depend on the degree to which that hypothesis is justified by facts (*op. cit.*, p. 580). According to him, a common source of divergence between the type of human behaviour that stems from the `rationality of the observer´ and the `rationality in the observed´ is the existence of a multiplicity of ends in actors´ minds. To the extent that the goals of actors are also an element of P-S, the adoption of a subjectivist interpretation of *RP* will require that the theoretician *understands* the goals of actors without this necessarily implying that she shares them. Now, the relevance of his notion of `subjective rationality´ emerges clearly in those cases where the situational model constructed on the basis of the `rationality of the observer´ *does not fit the facts*. As Schumpeter notes, in such cases the task of the theoretician is to explain the reasons for the *discrepancy* between the `rationality of the observer´ and the `rationality in the observed´ (*op. cit.*, p. 586). In turn, this will require an effort by the former to adopt the point of view of the `observed´:

`Understanding an end and judging rationality of means often requires that the analyst "puts himself" into places very far distant from his time and social location. Sometimes he has to transplant himself into another cultural world´ (Schumpeter, 1984, p. 583).

These ideas on the methodology of the social sciences were originally written by Schumpeter *circa* 1940 for a Harvard discussion group on rationality which included Parsons, Leontief, and Sweezy. The manuscript remained unpublished for more than 40 years until Professor Loring Allen of the University of Missouri in St. Louis found it among the papers of Schumpeter in the Harvard University archives. It was published posthumously in 1984 at the *Journal of Institutional and Theoretical Economics*. We believe this manuscript contains some intuitions that exhibit a high degree of affinity with Popper´s notions of *RPo* and *RPs* (Popper, 1994). However, we should like to note that it is Hayek´s *Economica* essay `Economics & Knowledge´ (Hayek, 1948[1937]) where the distinction between `objective´ and `subjective´ rationality was first formulated. In that essay, Hayek criticises equilibrium economic theory for making an illegitimate use of the concept of `data´ possessed by economic agents as well as for the methodological confusion thus created:

`But this does not solve the question whether the facts referred to are supposed to be given to the observing economist, or to the persons whose actions he wants to explain, and if to the latter, whether it is assumed that the same facts are known to all the different persons in the system, or whether the "data" for the different persons may be different… There seems to be no possible doubt that these two concepts of "data", on the one hand

in the sense of the objective real facts, as the observing economist is supposed to know them, and on the other hand in the subjective sense, as things known to the persons whose behaviour we try to explain are really *fundamentally different and ought to be kept carefully apart*. And, as we shall see, the question why the data in the subjective sense of the term should ever come to correspond to the objective data is one of the main problems we have to answer´ (Hayek, 1948[1937], p. 39, emphasis added).

We know that Popper had read Hayek´s 1937 paper in *Economica* and, indeed, he refers to it as the key source of his understanding of the core of economics (Popper, 1994, p. 181, footnote 1). According to Popper, it was Hayek´s exposition of the `logic of choice´ in that paper that led him to the formulation of the `logic of the situation´ as embracing both the `logic of choice´ and the `logic of historical P-S´. Yet, Popper does not refer explicitly to Hayek´s distinction between subjective and objective data. That said, it is very likely that Popper´s recognition, later on, of a distinction between *RPo* and *RPs* is related to his acquaintance with Hayek´s *Economica* essay. In the following sections we will explore in some detail the relation between these two concepts as well as their relation to PTKL *from the point of view of the agents* that are the object of the modelling exercise performed by the theoretician.

4.1.1. PTKL *versus* the `subjectivist´ version of *SA*

Let us focus on the relation between RPs and PTKL. To be sure, RPs constitutes a minimal requirement for rationality in that agents´ behaviour only has to be adequate or appropriate to the P-S as they see it. This implies that, as in the Austrian School of von

Mises, Hayek, and Schumpeter, rationality is associated to behaviour that is goal-directed or purposive.21 This type of rationality is sometimes denoted as instrumental in the sense that reason becomes an instrument to reach a certain goal, e.g. an increase in pleasure.22 Although there are significant methodological differences among members of the Austrian School of economics, they all viewed economics as part of a science of human action whose core is `to be found in the unique property possessed by human beings of engaging in operations designed to attain a state of affairs that is preferred to that which has hitherto prevailed´ (Kirzner, 1976, p. 148). What is crucial in our context is that the Austrian School´s conception of rationality is subjective in the sense of being an a priori assumption about human behaviour. There are two sources of subjectivity. First, there is the subjectivity of actors´ ends or, as von Mises puts it:

`Nobody else than the individual himself can decide what satisfies him better and what less... There is no such thing as an absolute state of satisfaction or happiness irrespective of the desires of the individual concerned´ (von Mises, 1944, p. 533).

Second, there is the subjectivity of knowledge itself in the social sciences. As long argued in Hayek (1943), it is only in the social sciences that our *interpretation* of a situation no matter whether it is right or wrong becomes an integral part of the situation thus affecting subsequent developments. Further, and to the extent that we understand the surrounding world via the 'internal models' we create, our understanding of the world will affect our decisions and, in this way, it may affect the world itself. Hayek (*op. cit*.) illustrates this theme by explaining the purposive nature of human action. As he explains, just as we cannot speak of the objective properties of a tool without saying something

about the purpose for which the tool is used so we cannot speak of social institutions objectively. Laws and economic institutions cannot be known apart from the intentions of the individuals who use them. In the field of economics, for instance, the value of money depends on the opinions of individuals who use it rather than on any inherent property of it. As we argue below, Hayek´s ideas on the methodology of the social sciences seem to have been ignored by most commentators of Popper´s work in that field.

Members of the Austrian School of economics like von Mises or Hayek adopted the `praxeological approach´ which consists of a theory of human action based on a set of self-evidently true *a priori* axioms on behaviour which, in turn, yields conclusions which are true regardless of time and place. However, the axioms of praxeology are not arbitrary like, for instance, those of mathematics. Rather, Austrian economists maintain that these axioms are already given to us in our minds and that, through the exercise of `introspection´ or `verstehen´, which consists of understanding the functioning of our minds, we have the possibility of understanding the behaviour of others. That said, the extreme subjectivism of the Austrian School of Economics leads to the conclusion that there is no possibility of acquiring knowledge about any social phenomena other than through `introspection´. Further, the notion of rationality proposed by von Mises (1944) as *purposive* behaviour may preclude the generation of predictions which can be subject to empirical tests. This is because the hypotheses about social phenomena derived from self-evident axioms may be close to being true but they may also possess little empirical content. This problem is addressed in Popper (1963, pp. 217-19) who makes it clear that science characterises as preferable `the theory which tells us more; that is to say, the theory which contains the greater amount of empirical information or content´. In other words, the empirical content of a theory increases with the increasing *improbability* of it being true or else with its increasing exposure to falsification. He uses the example of

meteorological forecasts; a forecast according to which in some unspecified time in the future it will rain has, as he explains, a high probability of being true yet it has virtually no empirical content, whereas a forecast which specifies the date and the time it is likely to rain has a high degree of empirical content yet it is quite likely to be false. Likewise, predictions derived from general or self-evident axioms on human behaviour like the ones of praxeology — which amount to stating little more than all human behaviour is purposeful — have a high probability of being true yet they have little empirical content because they are not falsifiable. By contrast, the subjectivist version of *SA* proposed by Popper (1985) is not subject to the previous criticism since, in addition to incorporating all the relevant elements of P-S — including the physical and social constrains and the knowledge and information possessed by agents — it also posits that actors´ behaviour is `adequate´ to P-S *as they see it*. The requirement that actors´ behaviour is `adequate´ — in addition to being purposeful or goal-oriented — implies, in turn, that the empirical content of theories constructed upon *RPs exceeds* the empirical content of theories about human behaviour derived from praxeology.

Let us distinguish between 'means-rationality', 'beliefs-rationality', and 'ends-rationality' (Hamlin, 1986). 'Means-rationality' implies the *correctness* of one´s actions *given* one´s desires and beliefs regardless of whether the latter are right. Therefore, as a minimum 'means-rationality' implies *consistency* of choice by agents. For instance, in standard consumer theory, 'means-rationality' is characterized by consistency in the preferences of households or *transitivity*: if an agent prefers *a* to *b* and *b* to *c*, then *a* must also be preferred to *c*. [23] The further requirements that are usually imposed, i.e., that individuals´ preferences exhibit both 'completeness' and 'continuity', are not ones of 'means-rationality' but rather of the optimization methods through which economists seek to represent individual preferences by a 'utility function'. Thus, when economists

speak of `rational´ agents what they usually have in mind is that their choices have to be, at least, consistent with one another.

Next, following Hamlin (*op. cit*), 'beliefs-rationality' implies that an individual´s (subjective) model of the surrounding world represents a good enough approximation to reality. Similarly, Bicchieri (1992) defines 'epistemic' rationality as a characteristic of beliefs that consists in their being *correct* given the evidence that is available to agents. Admittedly, both definitions are imprecise. For instance, in the former case it is unclear what a `good enough´ approximation to reality is. Similarly, in the latter case, and to the extent that the evidence that is available to every individual is necessarily limited, one might argue that agents´ beliefs are *always* correct given the evidence that is available to them. Thus, and for the purposes of this essay, let us characterise `beliefs-rationality´ as implying that agents´ subjective view of P-S coincides, with the theoretician´s view of it. In turn, `beliefs-rationality´ may be satisfied under any of the following scenarios: (i) agents exhibit `perfect foresight´ in which case their view of P-S always coincides with the theoretician´s and, hence, they do not make mistakes provided they also exhibit `means-rationality´, (ii) agents exhibit `rational expectations´ (Muth, 1961) — which is the stochastic version of the `perfect foresight´ case — in which case agents´ view of P-S is such that, provided they also exhibit `means-rationality´, their mistakes will only be random, and (iii) agents behave according to Subjective Expected Utility (SEU) theory (Savage, 1954) in which case the possibility that agents´ mistakes are systematic cannot be ruled out *a priori*. To be sure, in SEU theory it is assumed that the decision-maker knows the set of all possible consequences that will obtain from the adoption of every conceivable `course of action´ or `act´ even though she does not know which particular `state of the world´ will realize and, hence, *which* specific consequence will follow if she adopts a given `act´. The only source of uncertainty in SEU theory stems from the fact

that the numerical probabilities that decision-makers attach to every conceivable `state of the world´ are subjective and, hence, liable to error. However, and crucially, there is no presumption in the theory that the theoretician knows the `true´ value of the probabilities corresponding to each `state of the world´ so that the fulfilment of `belief-rationality´ in this case stems from the fact that it is assumed that both decision-makers and the theoretician know the precise consequences of adopting every possible `course of action´ yet they do not know *which* `state of the world´ will be realized.[24] Of course, if Savage´s version of SEU theory is supplemented by the notion that decision-makers´ (subjective) probabilities are correct `on average´ then this scenario becomes equivalent to the `rational expectations´ case.

Finally, 'ends-rationality' means that behaviour is *purposeful* or oriented to the achievement of a goal and, hence, not the result of chance (Hamlin, *op. cit*.). In the context of neoclassical economics 'ends-rationality' is usually associated to the pursuit of *self-interest*. To be sure, this has been the case at least since Edgeworth who, in his 1881 *Mathematical Psychics*, stated that 'the first principle of Economics is that every agent is actuated only by self-interest'. The typical statement of mainstream economics methodology prior to the emergence of 'neoclassical' economics in the second half of the nineteenth century, is in Mill (1967[1836]). Be that as it may, Twomey (1998, p. 435) claims that the clearest statements of this tradition were already present in the work of Bentham and Hobbes. In particular, he argues that Bentham (1907[1789]) first formulated the notion that agents seek to maximise pleasure whereas Hobbes (1985[1651]) provided a statement of egoism according to which individuals *always* seek their own greatest good. Therefore, we may characterise the `praxeologic´ models of the Austrian school of economics as implying `ends-rationality´ and the situational models based upon *RPs* as implying both 'means-rationality' and 'ends-rationality' but *not* 'beliefs-rationality'. As

we will argue below, each of these types of models implies a different division line between `rational´ and `irrational´ behaviour.

The absence of 'beliefs-rationality' in models based on *RPs* implies that agents´ beliefs may be wrong, i.e., agents are fallible and, crucially, that such mistakes may be systematic. Specifically, agents may perceive the physical and social constraints they face erroneously or may simply possess wrong information. Therefore, the adoption of *RPs* implies that agents´ beliefs may be (systematically) wrong and understanding their behaviour (including their mistakes) will require the construction of a situational model of the typical P-S *as seen by agents*. It follows from this that the adoption of *RPs* is *a priori* compatible with PTKL since the agents in the typical model can make mistakes stemming from their wrong beliefs. Yet, the notion of adequate behaviour according to PTKL also implies, as we noted above, the requirement that agents `learn´ from their past mistakes, i.e., they do not repeat them. As a result of it, full compatibility of *RPs* with PTKL would require that the theoretician recognizes that agents do not repeat their mistakes in the future. As we will see below, this feature of PTKL does create a tension with *RPs* when the purpose of the modelling exercise is to make predictions.[25] We may also add that, if *RPs* is adopted, the point of view of the theoretician *vis-à-vis* the actors becomes analogous to the position of participants in the `Beauty Contests´ that were so popular in the British tabloids in the 1930s and that were metaphorically captured by Keynes in his *General Theory* to explain the formation of the prices of financial assets (Keynes, 1936, p. 156). In `Beauty Contests´, what participants were supposed to do in order to win the prize was not so much to identify — among the photos of beautiful ladies portrayed in a panel — the lady they believed to be the most beautiful one but to `guess´ the photo of the lady they believed other participants would select as the most beautiful one. Similarly, we will argue below that, if *RPs* is adopted, the theoretician will seek to

reconstruct P-S not as she sees it herself but *the way she thinks agents see it*. In a `Beauty Contest´, what participants were supposed to do in order to win is not to identify — among the photos of beautiful ladies portrayed in a panel — the lady they believe to be the most beautiful one but to `guess´ the photo of the lady they believe other participants will select as the most beautiful one. In a similar fashion, we will argue that, if *RPs* is adopted, the theoretician will seek to reconstruct P-S not as she sees it herself but *the way she thinks agents see it*.

Next, and crucially, to the extent a *discrepancy* exists between the theoretician´s (objective) view of P-S and his conjecture about agents´ view of P-S, the generation of predictions will require making the crucial assumption that *such a discrepancy and the `situational factors´ that warrant it exhibit a high degree of stability over time*. In turn, the former implies that the `null hypothesis´ in empirical tests applied on a situational model which adopts *RPs* is that agents´ view of P-S is, at least partially, *wrong* whereas the alternative hypothesis is that the theoretician´s (objective) view of P-S is correct. If such discrepancy were to disappear over time for some reason (e.g., learning by agents), the predictions derived from it would be equivalent to the predictions generated if *RPo* were adopted. Thus, we disagree with Vanberg (2002, p. 12) when he argues that the `subjectivist´ *RP* poses a testability problem vis-à-vis the `objectivist´ *RP*. Specifically, situational models that adopt *RPs* can generate predictions albeit, as we argued above, their generation implicitly implies adopting the assumption that agents´ view of P-S *will remain constant in the future*. Unless the theoretician does so, the models´ testability will be seriously weakened. This is because, in the wake of an unfavourable empirical test, the theoretician may try to circumvent its refutation by arguing that the adverse result of the empirical test was due, for instance, to an (unpredictable) *change* in agents´ view of P-S. Thus, a *sine qua non* condition for potential refutability in this case, i.e., for the

model to possess `empirical content´, is that agents´ view of P-S is assumed to remain constant over time or, else, that agents follow a constant pattern of behaviour in spite of the mistakes such behaviour may bring about. However, this assumption creates some tension with PTKL since, according to the latter, agents tend to purge their wrong beliefs over time.

Now, we have argued above that, if *RPs* is adopted, it is implicitly assumed that (i) there is a discrepancy between the theoretician´s view of P-S and agents´ view of P-S, and (ii) that the former persists over time. As we argued above, this implies (under the null hypothesis) that the agents whose behaviour the theoretician seeks to capture in the situational model *do not revise their wrong beliefs* which runs counter to PTKL. The discrepancy alluded to above is between two different conjectures: (i) the theoretician´s (objective) view of P-S, and (ii) her view of agents´ view of P-S. According to Popper (1994, p. 178), the latter is always part of the former since the theoretician can only understand agents´ view of P-S if she reconstructs a *wider* view of P-S than their own. Specifically, if we adopt Popper´s interpretation of *RPs*, what the theoretician subjects to empirical test is the hypothesis that agents *systematically* fail to perceive the `true´ P-S and, under the null hypothesis, this implies that agents´ view of P-S will likely be disappointed *if* the theoretician´s view of P-S is correct. However, as we have argued above, the systematic disappointment of beliefs will come about because it is implicitly assumed that agents *do not `learn´ from their mistakes,* i.e., they tend to repeat mistakes all over. However, the above-mentioned tension between PTKL and *RPs* does not arise if the main purpose of constructing a situational model is to *explain the past* (e.g., historical interpretation) rather than to generate predictions. This is because in the former case the theoretician need not be concerned about the persistence into the future of a discrepancy between agents´ view of P-S and her `objective´ view of P-S. In short, *RPs*

is more problematic than Popper admits if the purpose of constructing a situational model is to generate predictions.[26]

To finish off this section, a clear example of this tension between PTKL and *RPs* is the Keynesian-type business cycle theory proposed by Minsky (1975) which is based on overoptimistic expectations of economic agents about their ability to honour future cash commitments which result from their inherited liability structure. According to Minsky´s `financial instability hypothesis´ (*op. cit*.), market economies are intrinsically unstable owing to the fact that economic agents become *systematically* overoptimistic during the upswing which makes them take on an excessive amount of debt and this eventually triggers off an asset price deflation and a subsequent financial crisis that precipitates the economy into a downswing. In other words, Minsky´s theory posits that, as memories from the last financial crisis fade out, agents will tend to underestimate the risk implied by the increase in the level of real indebtedness so that the upswing ends up when an external factor, e.g., an increase in interest rates, leads to an initial decrease in the price of real and financial assets which then brings about a reassessment of liability structures and, finally, leads to an asset price deflation. In other words, Minsky´s theory is a clear example of business cycle theory where (i): there is a discrepancy between the theoretician´s view of P-S and her view of agents´ view of P-S, and (ii) it is (implicitly) assumed that agents tend to repeat their past mistakes so the same phenomenon (i.e., business cycles), occurs recurrently and inevitably.

4.1.2. PTKL *versus* the `objectivist´ version of *SA*

According to Hands (1992, p. 28), `it is easy to see that situational analysis is the method of microeconomics (and of any macroeconomics based on micro foundations)´.

Indeed, Popper recognizes that his source of inspiration for *SA* is the methodology of neoclassical microeconomics (Popper, 1966[1943*a*], p. 97; 1944-45, p. 82; (2002[1976a], p. 93; 1976*b*, p. 117f).[27] However, on those few occasions when Popper makes it clear that he intends to extend the methodology of neoclassical economics to the rest of the social sciences he seems to have in mind the `objectivist´ *SA*.[28] That said, some commentators have noticed that the rationality requirements are more demanding if *RPo* rather than *RPs* is adopted (Latsis, 1983; Farmer, 1998; Oakley, 1999; Vanberg, 2002). Notably, and in addition to the fulfilment of both 'means-rationality' and 'ends-rationality' *RPo* also implies fulfilment of 'beliefs-rationality'. In standard consumer theory the presence of `perfect foresight´ by agents is pervasive so the fulfilment of `beliefs-rationality´ is unambiguous. Similarly, in macro-models which incorporate the `Rational Expectations Hypothesis´ (REH), the fulfilment of `beliefs-rationality´ implies that agents´ forecast errors are random. However, the presence of `beliefs-rationality´ in SEU theory requires some clarification.

In SEU theory, a `prospect´ is defined as an *exhaustive* list of `consequences´ stemming from different courses of action or `acts´ under different `states of the world´. Consequences are mutually exclusive possibilities. The decision-maker is assumed to be able to attach a (subjective) numerical probability to each of the consequences such that probabilities sum up to unity. The preferences of an individual are defined over the set of all conceivable `acts´. In other words, an individual must have a preference ordering that is *complete* over the set of conceivable `acts´ if she is to maximize expected utility. The evaluation by an individual of a given `act´ in the face of uncertainty involves both her preferences for the set of possible consequences and her (subjective) evaluation of the relative likelihood of each `state of the world´ actually being realized. Further, the preference structure is depicted axiomatically thereby allowing the theorist to transform,

by virtue of a von-Neumann-Morgenstern utility function, the preference ordering of any individual into a numerical value. Savage (1954, p. 30) interprets probability as `degrees of conviction´.[29] More generally, we can think of probability as an individual´s `degree of belief in the likely realization of an event´. Be that as it may, Savage (*op. cit.*, pp. 82-86) makes it clear that the practical applicability of SEU theory is restricted to `small worlds´ in which its axioms do apply.[30] Yet, there is by now plenty of evidence in the field of both experimental and behavioural economics showing that the axioms of SEU theory are often violated *even* in so-called `small worlds´. In any case, we believe that the subjective nature of probability in SEU theory does not preclude it from being part of the family of models characterised by *RPo*, the reason being that an axiomatic structure is *imposed* upon agents´ behaviour by assuming that a complete preference ordering of an (exhaustive) list of `acts´ based on subjective probability evaluations can be derived. In other words, the joint assumption that agents *know* the consequences of all possible `acts´ under all conceivable `states of the world´ — even though they do not know *which* `state of the world´ will be realized — and that they attach numerical probabilities to each `state of the world´ confers SEU theory a logical structure that places it within the family of models embodying *RPo* by fulfilling `means-rationality´, `ends-rationality´, and `beliefs-rationality´.

The feature that distinguishes SEU theory from the other scenarios identified above in which the assumption of `beliefs-rationality´ is satisfied is the possibility that agents´ mistakes are *systematic*. Let us explore this feature in some detail. First, Savage (1954, p. 20) admits that the axioms of SEU theory are preferably to be interpreted as principles that a rational individual *ought* to follow (i.e., as a prescriptive hypothesis) rather than as a descriptive hypothesis about *how* people actually choose among a set of possible `courses of action´.[31] Arguably, the possibility that agents´ errors are systematic

due to the subjective (and potentially wrong) nature of probability evaluations leads him to admit that the descriptive and, hence, the predictive power of SEU theory is rather limited unless it is supplemented by an additional assumption about the characteristics of subjective probability evaluations of individuals. Second, his recognition that SEU theory only applies to `small worlds´ also suggests that, even if we accept that the main usefulness of the theory is normative, the latter is valid solely in an, relatively small, section of reality. We believe that Savage´s admission that: (i) SEU theory only applies to `small worlds´, and (ii) that its main usefulness is *normative* lends support to our notion that the former is a particular implementation of *RPo* in which agents do exhibit `means-rationality´, `ends-rationality´, and `beliefs-rationality´. In particular, fulfilment of `means-rationality´ and `ends-rationality´ stems from compliance with the axioms of SEU theory and the (expected) utility maximizing behaviour of individuals respectively whereas fulfilment of `beliefs-rationality´ stems from the fact that, although probability is subjective and, hence, liable to error, agents know the set of consequences associated to any conceivable `act´ under every possible `state of the world´.

Now, the combination of the three types of rationality alluded to above yields a type of rationality known as 'substantive rationality' (*SR*) (Simon, 1976). *SR* is a type of rationality that is concerned exclusively with the *consequences* or outcomes of rational choice. In particular, Simon (1976, p. 130) denotes behaviour as being substantively rational 'when it is appropriate to the achievement of given goals within the limits imposed by given conditions and constraints'. Thus, as with classical decision theory, the interest lies not so much in *how* decisions are made but in *what* decisions are made. In short, *SR* constitutes a special type of application of *RPo* in that, in addition to both consistent and purposeful behaviour, it is assumed that *agents´ beliefs are correct on average*. As we have noted above, it is this approach to human rationality that lies at the

core of standard consumption theory, SEU theory and macro-models that incorporate REH. Indeed, some scholars have argued that *RPo* is the principle that underlies the methodology of mainstream economics (Farmer, 1998; Oakley, 1999).[32]

4.1.2.1. Rationalizing the `objectivist´ version of *SA*

Now, one can rationalize *RPo* as a *methodological decision* according to which the theoretician assumes beforehand that the mistakes made by agents (by `mistakes´ we mean decisions that are adequate from the point of view of P-S as seen by the agents but *inadequate* from the viewpoint of the P-S as seen by the theoretician) *are declared to be less interesting for the purpose of understanding agents´ behaviour and, especially, for generating predictions than the modelling mistakes made by the theoretician.* In other words, a rationale for *RPo* is that, although agents´ mistakes cannot be ruled out *a priori* — so *RPo* would be compatible with fallibility — nevertheless the theoretician *chooses to ignore the former for methodological reasons*. What are these reasons? First, that the theoretician *gains little*, if anything, by learning that agents make mistakes because (i) she already knows it and, more importantly, (ii) that the nature of the mistakes agents make is likely to change over time in an *unpredictable* way and so learning about them is of little help for the purpose of generating predictions. To be sure, learning about the mistakes agents have made in the past may be helpful if we had the assurance that the *same* mistakes (i.e., mistakes triggered off by the very same factors) will be repeated in the future. As we explained above, it is this scenario that may justify the adoption of *RPs*. However, *if such condition is not satisfied* there is arguably very little we can learn from the mistakes made by agents in the past other than for the purpose of historical analysis. A second reason for adopting this methodological decision would be that, if it were not adopted then in the wake of an erroneous prediction the theoretician might be tempted to

sidestep its falsification by arguing that agents´ beliefs and decisions were, on a particular episode — the one subject to the test — different from what one would have `objectively´ expected and to utilize this apparent `anomaly´ as a justification for the (adverse) result of the empirical test. By contrast, if *RPo* is adopted then the *onus of proof* will inescapably rest on the theorist´ view of P-S. In other words, the adoption of *RPo* implies that the theorist rules out the possibility that a wrong prediction generated by the model can be ascribed to agents´ mistakes and, hence, she forces herself to avoid resorting to immunizing strategies to prevent the refutation of the model.

According to the rationale for *RPo* we have suggested above, ascribing a role to agents´ errors in the situational model (i.e., to discrepancies between their view of P-S and the theoreticians´ view of P-S) will prevent us from generating predictions unless agents´ errors are predictable. Yet, one possible reason why agents´ errors may actually be unpredictable is that agents may `learn´ from their past mistakes so that their future mistakes will tend to *differ* from previous ones. Thus, in order to generate predictions, a hypothesis which assumes that agent´s view of P-S does not coincide with the theorist´s view of it (i.e., *RPs*) will need to be coupled to an additional assumption according to which agents´ errors tend to persist over time and, hence, are *predictable*. However, to the extent that this assumption implies that agents do not `learn´ from their mistakes, it is in conflict with PTKL. To be sure, if agents `learn´ from their previous mistakes so that they do not repeat them, their future mistakes will tend to *differ* from their previous mistakes and, unless the range of potential mistakes is limited, their future mistakes will thus be unpredictable. By contrast, if the theoretician adopts *RPo* instead of *RPs*, this problem does not arise because there is no presumption that agents´ view of P-S differs from the theoretician´s.

As we will argue below, the adoption of *RPo* necessarily implies the *imposition* of the theoretician´s view of P-S upon agents. Does this mean that there is no tension between the `objectivist´ version of *SA* and PTKL? We believe not. Firstly, PTKL only implies that agents `learn´ by trial and error-elimination so behaviour that is `rational´ according to PTKL may not be appropriate to the `logic of the situation´ as faced by agents. To be sure, the errors that agents made in the past occurred in the environment that surrounded them at that time so that if the latter changes in an unpredictable way agents may make new (and different) errors and, hence, they may make decisions that are inappropriate to the `logic of the (new) situation´. Secondly, since the theoretician´s view of P-S does not necessarily coincide with agents´ view of it, further assumptions will need to be made to justify the *coincidence* of the theoreticians´ and agents´ views. These additional assumptions will be presented and discussed below. In any case, we may anticipate that these assumptions are problematic in the sense that if it is assumed that agents actually behave according to PTKL it is doubtful that their decisions will be appropriate to the `logic of the P-S´ as seen by the theoretician, even `on average´. In short, the nature of tension between PTKL and *RPs* differs from the nature of tension between PTKL and *RPo* in that, in the former case, tension stems from the fact that the adoption of *RPs* implies that agents may *repeat* the same mistakes they made in the past in a way hardly compatible with `learning´ by trial and error-elimination as posited in PTKL whereas, in the latter case, it is *assumed* that agents´ view of P-S coincides, at least on average, with the theoretician´s view of P-S so that their behaviour is always appropriate to the objective `logic of the P-S´ yet, as we argue below, the mechanisms by virtue of which such coincidence is justified can hardly be reconciled with PTKL.

4.1.2.2. How do the views of the agents and the theoretician tend to converge?

We suggested above that the adoption of *RPo* implies *de facto* the *imposition* of the theoreticians´ view of P-S upon agents´. However, social scientists do not usually see things this way. An example is mainstream economics where a number of devices have been suggested in the literature to (implicitly) justify the adoption of *RPo*. To be sure, such mechanisms are seemingly viewed by mainstream economists as reasons why they need not care about agents´ beliefs when reconstructing P-S because an `objective´ P-S can be said to exist `out there´ that is *sufficiently* independent of agents´ beliefs (and actions) so the theoretician can reconstruct P-S as she sees it. According to us, there are two devices through which the neglect of agents´ beliefs by the theoretician is normally justified: (i) the operation of the `law of large numbers´ in the social domain, and (ii) the presence of `learning´ by individuals. However, as we argue below, both mechanisms are problematic. Let us address the first mechanism. According to it, agents´ decisions may turn out to be objectively wrong in retrospect but nevertheless their *mistakes will tend to cancel each other out provided the number of individuals is large enough*. The former implies that the scope for fallibility at the aggregate level in this version of *RPo* is negligible since it is restricted to random mistakes associated to transitory factors. There is some textual evidence which suggests that several influential social scientists implicitly resort to this mechanism to justify the adoption of *RPo*. For instance, Nobel Laureate in Economics John Hicks (1956, p. 55) writes that `the preference hypothesis [in the context of neoclassical utility theory] only acquires a *prima facie* plausibility when it is applied to a statistical average´. More explicitly, Gibbard & Varian (1978) describe optimizing behaviour as capturing the 'central tendency' of rational economic behaviour or:

‘If deviations are random or more precisely, are not systematic, there might be good reason to have some faith in the conclusions of the [economic] model even though the assumptions, strictly interpreted, are implausible. Perhaps a case in point is the economist´s assumption of perfect optimizing behaviour. Of course, this assumption is strictly speaking, false, but, so long as errors in optimization are not systematic, this hypothesis may be useful in describing the “central tendency” of economic behaviour. Furthermore, in models where individual units´ behaviour is being aggregated, non-systematic errors may be expected to “wash out” in the process of aggregation’ (*op. cit.*, p. 670).

Likewise, it has been argued in the sphere of sociology that it is not necessary to claim that all agents optimize but, instead, that the tendency to optimize is the most important *non-idiosyncratic* factor at work so that the operation of a sort of ‘law of large numbers’ guarantees that optimizing behaviour dominates (Goldthorpe, 1998, p. 169). We believe this assumption (i.e., the `law of large numbers´) implicitly lies at the core of neoclassical economics where agents are modelled *as if* they were infallible – when they exhibit perfect foresight – or as if their mistakes were random (Muth, 1961). More specifically, this assumption is implied when the optimizing assumption is applied in modelling exercises. However, let us note that the `law of large numbers´ in the field of statistics assumes that the different trials of a stochastic process are: (i) independent and, crucially, (ii) have the *same* distribution so that, as the number of such trials tends to infinity, the probability distribution of a random variable *concentrates* around the finite expected value of each of the trials.

Now, it is unlikely that these conditions will be satisfied in the case of agents´ view of P-S. For one thing, there are likely to be significant interdependencies among

agents´ (subjective) view of P-S owing to the presence of conventional elements so that condition (i) is likely to be violated. Further, agents´ (subjective) view of P-S may differ substantially from others agents´ which also violates condition (ii). Thus, the conditions for reliance on the `law of large numbers´ as it exists in statistics for the purpose of providing a rationale for the coincidence, on average, between agents´ view of P-S and the theoreticians´ view of P-S are not warranted. Consequently, the adoption of *RPo* can solely be justified on *strict* methodological grounds. That said, we believe that some advocates of *RPo* assume that if the theorist´s view of P-S diverges significantly from agents´ view of P-S such discrepancy will tend to be eliminated over time by means of other mechanisms such as: (i) trial and error-elimination, and (ii) imitation of successful strategies by agents (Alchian, 1950). In other words, advocates of *RPo* might argue that the occurrence of learning at the individual level based on trial and error-elimination and/or the imitation of the successful strategies of others will make agents´ view of P-S eventually converge to the theoretician´s (objective) view of P-S so that, for the sake of analytical convenience, we may ´confidently assume that agents´ beliefs are correct `on average´. However, the imitation of successful strategies requires that some other agents have previously `learnt´ to perform some tasks adequately so that the presence of some kind of learning is a *sine qua non* condition for the imitation of successful strategies to allow other agents to make decisions that are appropriate to the `logic of the situation´. Therefore, for the sake of simplicity, we will leave the latter aside. In other words, the theoretician assumes in this case that the operation of a *negative* feedback mechanism whereby agents systematically revise their (wrong) beliefs until the latter coincide with the theoretician´s view of P-S justifies the adoption of the methodological decision to assume that agents´ beliefs are correct `on average´. In short, `learning´ is the second mechanism (additional to the `law of large numbers´) by virtue of which mainstream

economists implicitly justify the assumption that agents´ view of P-S coincides, at least `on average´, with the theorist´s.[33] For instance, Nobel Laureate R. Lucas characterizes the type of situations on which economic theory focuses as the *end-result* of an adaptive learning process:[34]

'Economics has tended to focus on situations in which the agent can be expected to "know" or to have learned the consequences of different actions so that his observed choices reveal stable features of his underlying preferences… Technically, I think of economics as studying decision rules that are steady states of some adaptive process, decision rules that are found to work over a range of situations and hence are no longer revised appreciably as more experience accumulates…' (Lucas, 1986, p. 218).

However, for this feedback mechanism to be *effective*, it is necessary that: (i) the former is fast and accurate enough, and (ii) P-S remains constant until the convergence process has been completed.[35] Yet, as Tversky & Kahneman (1986, p. 90) insist, such conditions rarely arise in the real world. In particular, the former can hardly be satisfied when agents make decisions in a changing environment in which it is hard to ascertain whether an observed outcome is a direct consequence of our decisions or a consequence of someone else´s decisions. Furthermore, Popper (1994, p. 4) insists that *no optimal state of adaptation* is ever reached by the application of the method of trial and error-elimination owing to: (i) the continuous *change* in the environmental situation, and (ii) agents´ inability to eliminate all their errors. That said, we believe that it is the reliance on the alleged efficacy of `learning´ at the individual level that makes some scientists implicitly assume that any negative result that occurs in the wake of empirical tests can be ascribed only (or mainly) to their *own* modelling mistakes (i.e., to their own failure to

capture the `objective´ P-S properly) rather than to agents´ mistakes. Thus, although *RPo* accounts for the presence of learning, it exhibits some clear differences with PTKL in that `learning´ at the individual level is unlikely to warrant that agents´ decisions will be adequate to the `logic of the situation´ as it is *seen by the theoretician*.

4.1.3. The dichotomy between `rational´ and `irrational´ behaviour

According to Simon (1965, p. 84), theoretical models based on 'substantively' rational individuals share a common framework characterised by: (i) a set of alternative courses of action that are available to the individual, (ii) knowledge that permits the individual to predict the *precise* consequences of choosing each course of action, and (iii) a criterion for determining *which* set of potential consequences she prefers. In such models, rationality is usually defined as 'the ability of actors to select that course of action which leads to the most preferred set of predicted consequences' (*op. cit*.).[36] Therefore, SR assumes that the surrounding environment is either known or knowable (i.e., the stochastic environment is stable), and that individuals have sufficient cognitive abilities to deal with a complex reality. SR is the type of rationality actors are assumed to exhibit in models that adopt strong versions of *RPo* such as the ones that prevail in mainstream economics. In particular, agents who exhibit SR must fulfil ends-rationality, means-rationality, and beliefs-rationality. This is depicted below in the third row of Table 1. In turn, this implies that behaviour that falls short of maximizing is deemed 'irrational' (Becker, 1962). More specifically, violation of `means-rationality´ and/or `beliefs-rationality´ is interpreted by mainstream economists as signalling `irrational´ behaviour. Likewise, Popper (1966[1943a]), p. 97) explains that `when we speak of "rational behaviour" or of "irrational behaviour" we mean behaviour which is, or which is not, in accordance with the logic of the situation´. Thus, Popper´s notion of rationality in the

context of *SA* bears, arguably, a strong resemblance to the notion of rationality in mainstream economics.

That said, we believe that the charge of `irrationality´ is a direct implication of the adoption of *RPo* and, particularly, of the *imposition* upon agents of the theoretician´s view of P-S. More specifically, we believe that the `irrationality´ charge that is applied to those agents who fail to maximize a given pre-specified objective function obeys ultimately to a failure to distinguish between the `rationality of the theoretician´ and the `rationality of agents´. For instance, the implicit assumption by mainstream economists that agents´ (subjective) view of P-S *coincides*, on average, with the theoretician´s view of P-S implies that all behaviour that falls short of the rationality standard ascribed to the theoretician is `irrational´. However, we believe that if agents´ view of P-S does not coincide with the theoretician´s then the former cannot be blamed for being `irrational´. In particular, an individual cannot be said to be `irrational´ if, for instance, her beliefs are wrong. Rather, as PTKL has it, we can only be said to be `irrational´ if we refuses to revise our (wrong) beliefs.

The counterpart to SR is the notion of 'procedural' rationality (PR). According to Simon (1976, p. 131), 'behavior is procedurally rational when it is the outcome of appropriate deliberation'.[37] PR shifts attention from the consequences of choice to the *process* of choice where the emphasis is placed in the presence of a decision process based on the use of simple heuristics or 'rules of thumb'.[38] PR can thus be characterised as the ability of actors to use simple heuristics that are *adequate* for a specific purpose. Reliance on simple heuristics to make decisions assumes that, most of the time, actors face situations characterised by (i) Knightian uncertainty (Knight, 1971[1921]), or (ii) where the 'optimal' solution is intractable. The former corresponds to scenarios where either we do not have an exhaustive list of potential consequences of a certain decision or else to

situations where, even if such a list were available, it is impossible to attach numerical probabilities to them. In turn, the latter corresponds to situations where there are insurmountable constraints on agents´ ability: (i) to identify optimal actions given a set of beliefs and desires, and (ii) to acquire the information relevant to the problem at hand. In contrast, SR implies that, in conditions of imperfect knowledge, agents make decisions by following Bayes's rule or by maximizing their expected utility, as in SEU theory.

Now, we believe PR captures properly the type of rationality implied by PTKL. As we noted above, the acquisition of knowledge in Popper's account runs parallel to the *process* of adaptation to a partly unknown (and changing) environment in which some of the errors made in the past by individuals are purged by virtue of a learning process that consists essentially of subjecting their hypotheses or conjectures to trial and error-elimination. Crucially, Popper emphasises that our adaptation to the surrounding environment is often successful and often unsuccessful. More specifically, some errors will *escape* and this possibility is one of the reasons our knowledge is always fallible (Popper, 1990, p. 47). In turn, Popper (1994, p. 4) insists that the systematic application of the method of trial and error-elimination will not result in an 'optimum' adaptation to the surrounding environment. Instead, and due to the partial elimination of errors, our adaptation to the environment is always *imperfect*. It follows from this that the observed states of adaptation *can never be the result of convergence to an optimum*. Should *all* errors be systematically purged and the environment be stable, the process of adaptation to the latter would eventually be perfect and only then could the state of adaptation be interpreted as the outcome of a convergence to an optimum. In this scenario, individuals would be fallible only to the extent that the changes in the environment cannot be fully anticipated. The learning method of trial and error-elimination consists of an adaptation mechanism whereby errors are eliminated and new hypotheses are subject to trial. As

some commentators note (Kerstenetzky, 2009; Lagueux, 2006), the watershed between `rational´ and `irrational´ behaviour in PTKL is marked by the *unwillingness of agents to correct their wrong beliefs* or, as it were, by the *incorrigibility* of their beliefs. This is clearly stated by Popper in the following quotation:

`The main distinction, I suggest, is that a healthy person´s beliefs are not incorrigible: a healthy person shows a certain readiness to correct his beliefs. He may do so only reluctantly, yet he is nevertheless ready to correct his views under the pressure of events, of the opinions held by others, and of critical arguments… the mentality of the man with definitely fixed views, the "committed" man, is akin to that of the madman… but *in so far as he is committed, he is not rational´* (Popper, 1985, p. 364; 1994, p. 180, emphasis added).

This suggests, as noted above, *that there are two different notions of rationality in Popper´s work*: (i) behaviour that is in accordance with the `logic of the situation´ (Popper, [, p. 97; 1944-45, sections 31 & 32), and (ii) willingness to revise one´s wrong beliefs (Popper, 1985, p. 364). We have argued above that there is a certain tension between these two notions of rationality when the relation is approached from the standpoint of the `rationality of the agents´ and we argue below that there is no such tension when the relation is approached instead from the standpoint of the `rationality of the theoretician´.

Next, unlike most models based on *RPo*, both praxeology and models based on *RPs* imply that agents´ view of P-S may be wrong and, hence, that their decisions may turn out to be wrong *ex-post*. In the case of *RPs*, the theoretician is assumed to adopt the

viewpoint of actors and, thus, she is supposed to be able to distinguish between her own view of P-S (i.e., the `rationality of the observer´) and agents´ (subjective) view of P-S (i.e., the `rationality in the observed´). Furthermore, in the case of models based on *RPs*, agents are usually assumed to exhibit means-rationality (and `ends-rationality´). If so, then their mistakes can only be ascribed to wrong beliefs and not to an inconsistent or inadequate behaviour *given* the information available to them. However, as far as agents are concerned, there is no mechanism that ensures an adequate, let alone efficient, use of the available information. Rather, decision-making in a context of uncertainty can only be the result of a process of systematic deliberation. In other words, although we cannot rule out that a *given* correct decision is the outcome of sheer chance, it is much more likely that correct decisions will be the result of agents´ systematic deliberation given the knowledge they possess. In short, the type of rationality agents exhibit if *RPs* is adopted can be denoted as `procedural´. In turn, this implies that PTKL and *RPs* share the feature that agents are *fallible*, i.e., their beliefs may be wrong *ex-post*. By contrast, as we explained above, the scope for fallibility is negligible if *RPo* is adopted because although individuals make mistakes the latter are assumed to cancel out at the aggregate level.

Now, by adopting *RPs* the theoretician seeks to identify agents´ partially wrong beliefs and thus to explain their behaviour accordingly (i.e., by stressing the divergence of agents´ behaviour from what one would expect if their beliefs were `correct´). By contrast, it is quite unlikely that the theorist can provide a `rational´ reconstruction of agents´ apparently wrong behaviour by pointing instead to their *inadequate* behaviour given their *correct* beliefs (i.e., to absence of means-rationality). To be sure, to provide such an account of agents´ inadequate behaviour by appealing to the notion that some agents exhibit, for instance, inconsistent behaviour or weak willpower, implies a large element of psychologism and, hence, a great loss of inter-subjectivity, accountability, and

transparency in theoretical analysis. Further, such diagnose of inadequate behaviour as based on a violation of the assumption of `means-rationality´ may apply to some or even to a very small proportion of individuals but it will hardly apply to the majority of them which, according to us, precludes its use in the social sciences. This suggests that, even though it is *not* true that agents always make an adequate use of the information available to them and, to the extent that we accept as universally valid the assumption that agents´ behaviour is always goal-oriented, it is unclear whether we can speak of a dichotomy between `rational´ and `irrational´ behaviour in models which adopt *RPs*. By contrast, to the extent that agents only exhibit `ends-rationality´ in praxeologic models, we cannot speak of the existence of such a dichotomy in the latter. This is depicted in Table 1 below.

Approach	Beliefs	Rationality	Dichotomy
Praxeology	Right or wrong *ex-post*	Ends-rationality (instrumental)	No
Subjectivist SA	Right or wrong *ex-post*	Ends & means-rationality (procedural)	?
Objectivist SA	Right `on average´ (e.g., `perfect foresight´ and `rational expectations´ models)	Ends, means & beliefs-rationality (substantive)	Irrationality ⇒ failure of either means-rationality or beliefs-rationality (or both)
	Right or wrong *ex-post* (e.g., SEU theory)		
PTKL	Right or wrong *ex-post*	Ends-rationality (procedural)	Irrationality ⇒ incorrigibility

Table 1. Classification of approaches to rationality

Finally, let us note that it is the existence or otherwise of a dichotomy between `rational´ and `irrational´ behaviour in the context of *SA* that was the central object of an exchange between Nadeau (1993) and Lagueux (1993, 2010). In their contributions, which focus on the epistemological status of *RP*, Nadeau (1993) initially argued that *RPs* is the correct interpretation of Popper´s RP and subsequently argued that the former is a metaphysical statement and, as such, it is *a priori* true and irrefutable. By contrast, Lagueux (1993) held that *RP* should rather be interpreted as a methodological principle and, hence, that it is a false but approximately true principle in the sense there may be some instances where it does not hold. Nevertheless, he (as Popper does) believes that *RP*

is a sufficiently good approximation to the truth and thereby proposes a `statistical´ justification of *RP* in so far as he points out that `the rationality principle can be said to be "approximately true" only to the extent that it applies to a large number of cases´ (Lagueux, *op. cit*., p. 475). Both authors make use of Popper´s example of the `flustered driver´ to substantiate their arguments:

`For the rationality principle seems to me to be clearly false — even in its weakest zero formulation, which may be put like this: "Agents always act in a manner appropriate to the situation in which they find themselves."… In think one can see easily that this is not so. One has only to observe flustered drivers trying to get out of a traffic jam, or desperately trying to park their cars when there is hardly any parking space to be found, or none at all, in order to see that we do not always act in accordance with the rationality principle… Moreover, there are, obviously, vast personal differences, not only in knowledge and skill — these are part of the situation — but also in assessing or understanding a situation; and this means that *some people will act appropriately and others not´* (Popper, 1985, p. 361; 1994, p. 172, emphasis added).

Building on this, Lagueux comments:

`Suppose that we try to explain the [Popper´s] example of the flustered driver using the rationality principle as understood by Popper, that is to say, as a principle that *is empirical and false*. It is clear that, in such a case, the rationality principle *could* be held responsible for the failure of an explanatory theory that was supported by it´ (Lagueux, 1993, p. 475).

By contrast, Nadeau (1993) writes:

`It seems to me to be evident that the *RPs* is logically irrefutable, in exactly the same way that, for Popper, probabilistic assertions or metaphysical statements are irrefutable... My critique of Popper can be clarified further by a brief examination of his example of the flustered driver... Popper uses this example to falsify the *RP*, but how does this example work exactly? It merely makes it apparent that the real or objective situation is such that, in spite of the fact that there are no available parking spaces, the driver persists in trying to park his car. However, it is rather surprising that, in his analysis of the situation, Popper does not connect the irrationality of the driver to the *contradiction* between the information that the latter has at his disposal and the chosen course of action. For if the driver does not know or does not believe that the parking space where he is desperately trying to park his car is insufficient, then his behavior does *not* contradict the *RPs´* (Nadeau, 1993, pp. 461-62, emphasis added).

In an attempt to clarify this controversy Lagueux (2010, pp. 104f) notes that, the `flustered driver´ in Popper´s example, represents an atypical behaviour that cannot be excluded and that such behaviour can be said to be `irrational´. However, he recognizes that, `if we base our judgement *on the description alone*, we cannot be sure that the flustered driver´s action is really irrational´ (*op. cit*., p. 104). This is because we can *never* be certain that the agent sees P-S in a way that renders his behaviour appropriate. In other words, according to Lagueux (*op. cit*.) the same behaviour can be interpreted as either `rational´ or `irrational´ depending on the *observer´s viewpoint*. We thus believe that Lagueux (*op. cit*.) ultimately admits there is no objective or neutral way of deciding

whether someone exhibits `means-rationality´ whenever there is no assurance that her beliefs are correct. We conclude that, in the case of *RPs*, the dividing line or watershed between `rational´ and `irrational´ behaviour is unclear.

4.1.4. What version of *RP* is `more adequate´ in the social sciences?

The theoretician´s methodological decision to ignore agents´ mistakes if *RPo* is adopted implies, for the reasons expounded above, assuming that agents´ (subjective) view of P-S *converges* over time to, or else, does not diverge in a significant way from the theoretician´s view of P-S. More specifically, we argued above that the adoption of *RPo* implicitly implies the *imposition* upon agents of the theoretician´s view of P-S. By contrast, and despite the above-mentioned tension between PTKL and *RPs*, we believe that *a more natural strategy in the theoretical social sciences is to adopt agents´ view of P-S*. There are at least three reasons for this. First, as Popper (1972, p. 179) admits, both the theoreticians´ and agents´ view of P-S are *conjectured*. Second, and more important, there is Hayek´s notion that, unlike the facts of the natural sciences — which are largely independent of the theoretician´s viewpoint — the `facts´ of the social sciences are all *interpretations* (Hayek, 1943).[39] That is, according to Hayek, the concepts we use in the social sciences are not just abstractions like the ones used in the physical and natural sciences but they abstract from all the physical characteristics of the objects they refer to. Hayek (*op. cit*., p. 3) denotes the concepts we use in the theoretical social sciences as `teleological´ because, as he explains, such concepts can only be defined by postulating relations between three different terms: (i) a purpose, (ii) somebody who holds it, and (iii) an `object´ which the person in question thinks to be a suitable means to achieve that purpose. As he explains:

`We could say that all these objects are defined not in terms of their "real" properties but in terms of *opinions* people hold about them. In short, in the social sciences *the things are what people think they are*. Money is money, a word is a word, a cosmetic is a cosmetic, if and because somebody thinks they are. That this is not more obvious is due to the historical accident that in the world in which we live the knowledge of most people is approximately similar to our own… We are likely, for example, to think of the relationship between parent and child as an "objective" fact. But, when we use this concept in studying family life, what is relevant is not that *x* is the natural offspring of *y* but that either or both believe this to be the case (*op. cit*., emphasis added).

Third, Hayek (*op. cit*.) also argues that in the type of P-S analysed in the social sciences agents´ *interpretation* of P-S becomes an `integral´ part of the latter thereby affecting subsequent developments. In particular, to the extent that agents understand P-S via the internal models they create for that purpose, their understanding of the former will affect their decisions and, through this route, they *may affect P-S itself*. Let us use the example of `bank panics´ to illustrate this idea. The occurrence of a `bank panic´ in a private bank is not necessarily related to the *actual* liquidity position of the bank. Rather, the occurrence of a `bank panic´ is more likely to depend on its depositors´ view about the ability of the bank to cash their deposits on demand. If depositors have doubts about the ability of the bank to comply with its obligations when the former attempt to withdraw money from their accounts (and *regardless* of the `true´ liquidity position of the bank), a `bank panic´ will likely ensue and the bank will *actually* become illiquid. This is not to deny that depositors commonly take into account the `objective´ indicators related to the liquidity of the bank when evaluating the likelihood of the private bank going illiquid. Rather, our argument is that *what really matters as far as depositors´ decisions are*

concerned is not the (objective) information provided by the liquidity indicators but agents´ (subjective) evaluation of them. However, if the latter affects the former, then P-S is *not independent of depositors´ views* and, at least in this example, it is not sound to argue that there is an `objective´ P-S which is, in principle, knowable by the theoretician but *not* by the agents. This idea, we believe, is captured in the following comment by Hayek:

`Perhaps the relevant distinction comes out most clearly in the general and obvious statement that no *superior* knowledge the observer may possess about the object, but which is not possessed by the acting person, can help us in understanding the action in question´ (*op. cit*., emphasis added).

Unlike the presupposition by social scientists (and Popper) that the theoretician possesses a *wider* perspective of P-S than agents do, Hayek (*op. cit*.) suggests that, since P-S depends on agents´ interpretation no matter whether the latter is right or wrong, it follows that the theoretician does not stand in a privileged position to observe the `objective´ P-S. Thus, and for these reasons, we believe that the `natural´ strategy for the theoretician is to seek to capture P-S *as agents see it*. That is to say, if there is not an `objective´ P-S that is (fully) independent of agents´ views, then the theoretician has a better chance of understanding social phenomena if she adopts agents´ viewpoint. This is not to deny, however, that there may be some circumstances in which the theoretician may prefer instead, for methodological reasons, to adopt *RPo*. In particular, there may be circumstances where P-S may be sufficiently independent from agents´ beliefs as to make it convenient to adopt *RPo*. Be that as it may, the adoption of *RPo* will actually imply the *imposition* of the theoretician´s view of P-S upon agents or, as Schumpeter (1984) would

put it, the model will capture the `rationality of the theoretician´ instead of the `rationality in the observed´.

Now, the former discussion suggests that *RPo* represents a *limit* or *extreme* case of *SA*. In particular, we believe *RPo represents a limit case of SA based on the implicit assumption that P-S is (fully) independent of agents´ beliefs* and that, consequently, the theoretician can acquire `objective´ knowledge about P-S that is, somehow, *superior* to agents´. By contrast, Hayek´s ideas on the nature of the `facts´ of the social sciences imply that there are several elements of P-S such as the knowledge and information that agents possess and the social (and even physical) constraints their behaviour is subject to which depend on agents´ beliefs so the theoretician cannot claim to possess superior knowledge about them. In other words, the adoption of *RPo* could *a priori* be justified if the theoretician were able to acquire knowledge of P-S that is not available to agents but if, as Hayek (*op. cit*.) suggests, this is not the case, that is, if P-S consists, at least partly, of agents´ beliefs, if follows that *RPo* constitutes a *limit* case whose adoption implies *imposing* the theoreticians´ (allegedly superior) view of P-S on agents.

4.1.5. A reformulation of `situational analysis´

Popper apparently ignores both Hayek´s ideas about the peculiar `facts´ of the social sciences and his own ideas about indeterminism in the natural sciences in his discussion of *SA*. For instance, in his most detailed presentation of *SA* (Popper, 1994, p. 183, note 19), he argues that, if *RPo* is adopted, then the theorist reconstructs P-S *as it actually is* whereas, if *RPs* is adopted, she reconstructs P-S *as agents actually see it*. Yet, the way this is expressed by Popper is somewhat ambiguous since, in his attempt to clarify this issue, he seems to refer only to historical interpretation and, thus, it is unclear whether

the distinction he draws between *RPo* and *RPs* also applies to the other social sciences. Hereafter, we assume that it does but we should like to make it clear that this is our interpretation. Specifically, he writes that:

`It seems to me now that there are at least three senses of `rationality´ (and, accordingly, of the `rationality principle´), all objective, yet differing with regard to the objectivity of the situation in which the agent is acting: (1) *The situation as it actually was* — the objective situation which the historian tries to reconstruct. Part of this objective situation is (2) *The situation as the agent actually saw it*. But I suggest that there is a third sense intermediate between (1) and (2): (3) *The situation as the agent could* (within the objective situation) *have seen it*, and perhaps ought to have seen it´ (Popper, 1994, p. 183, footnote 19).

The previous quotation highlights that, when drawing a distinction between *RPo* and *RPs*, Popper assumes that the theoretician possesses knowledge that is superior to agents´. Otherwise, he could not have defined the third and intermediate sense of *RP* as one in which P-S is *`as the agent could* (within the objective situation) *have seen it*, and perhaps ought to have seen it´ (*op. cit*.). That is, the explicit reference by Popper to the P-S `as the agents *could* and perhaps *ought to* have seen it´ logically implies that he is assuming implicitly that there exists an `objective´ P-S and that the theoretician is in a *better* position than agents to observe it. Thus, we think Popper fails to take on board Hayek´s ideas about the peculiar `facts´ of the social sciences as well as his own views about indeterminism in the natural sciences. This takes us to our following claim. We believe that, arguably, the real difference between *RPo* and *RPs* is not that in the former

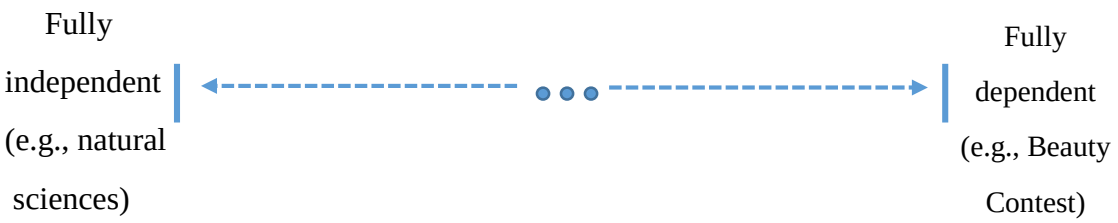
the theorist reconstructs P-S *as it actually is* (even if understood as being conjectural) whereas in the latter she reconstructs it *as agents actually see it* (also understood in a conjectural way) but, instead, that in the former the theorist reconstructs P-S *as she sees it* whereas, in the latter, the theorist reconstructs it *as she believes agents see it*. This suggests that the difference between *RPo* and *RPs* in this reformulated framework is not the `objectivity´ of the approach — because both the theorist´s and agents´ view of P-S are subjective — but the *degree* in which the subjectivity of the theorist manifests itself; in the case of *RPo* the implied subjectivity of the theorist is of a `first degree´ because it is her (direct) view of P-S that is at stake whereas in the case of *RPs* the subjectivity of the theorist is of a `second degree´ because in that case it is her view about agents´ view of P-S.

Let us put it another way, if *RPo* is adopted the theorist reconstructs P-S *as she sees it* and, consequently, different theorists may reconstruct it in different ways (as it is commonly the case in the theoretical social sciences). Likewise, if *RPs* is adopted the theorist will reconstruct P-S *as she thinks agents see it*. Again, in this second scenario, different theorists may have very different views about *how* agents see P-S and, hence, may produce different theories about the same phenomenon. The reason is that, *if P-S is, at least partly, as agents see it, then it follows that Popper´s distinction between P-S `as it actually is´ and P-S `as agents see it´ makes little sense*. Indeed, we may add that in the extreme case in which P-S fully coincides with agents´ views (e.g., the Beauty Contest metaphor), *P-S is as agents see it* and, therefore, *RPo* and *RPs* would become equivalent if they were defined as Popper does.

Now, this suggests that, as we show in Chart 1 below, *SA* exhibits a spectrum of potential scenarios according to the *degree of independence* of P-S from agents´ beliefs. At one extreme of the spectrum there are those cases characterized by *full* coincidence of

P-S with agents´ beliefs or, as it were, by the absence of elements in P-S whose properties can be said to be (fully) independent from agents´ beliefs. Again, an example of this scenario is Keynes´ Beauty Contest metaphor in which there are no objective `facts´ the theoretician can observe because agents´ opinions about the (relative) beauty of the ladies portrayed in the photos are subjective. At the other extreme of the spectrum there is the typical scenario in the natural and physical sciences in which the `facts´ can be said to be *fully* independent from the observers´ viewpoint. Hayek´s (1943) suggests that in the social sciences there is no such scenario since P-S is *never* independent from agents´ beliefs and, hence, if we take Hayek´s ideas on board the former can be said to be a *limit* or extreme case. Between these two extreme cases there is a spectrum of potential P-S characterized by different positive degrees of dependence of P-S upon agents´ beliefs so that the lower the degree of dependence the closer the scenario will be to *RPo*. Finally, let us add that the case known as `self-fulfilling´ expectations (Merton, 1948) would correspond to the case in which agents´ beliefs, no matter whether they are right or wrong, bring about a change in P-S so that the latter eventually *converges* to the former. In our framework, this could only occur if P-S were at one end of the above-mentioned spectrum; the one characterised by coincidence of P-S with agents´ beliefs.

Chart 1. Spectrum of scenarios according to the degree of `independence´ of P-S from agents´ beliefs



Let us finish off this section by adding that the former discussion highlights that, if *RPo* is adopted, the null hypothesis in an empirical test is not that the theorist´s view of the typical P-S is correct, as it is usually believed, but rather that *agents´ view of the typical P-S coincides, on average, with the theorist´s*. That is to say, to the extent that agents´ view of P-S is an integral part of the latter, only if there is a large coincidence (on average) between the agents´ and the theorist´ view of P-S will the hypothesis have a chance of withstanding the *onus of proof* when subject to an empirical test. In other words, an empirical test is not, if *RPo* is adopted, a contrast between the theorist´s view of P-S and the `objective´ facts, as is the case in the natural sciences. Rather, as we have argued above, the `facts´ of the social sciences are largely *interpretations* and, hence, rejection of the null hypothesis in this case does not imply that the theorist´s view of P-S does not capture the `objective´ facts appropriately since there are no `objective´ facts in the social sciences but, rather, that the theorist´s view of P-S does not coincide, on average, with agents´ view of P-S.

4.2. PTKL *versus SA*: the `rationality of the theoretician´

The last issue we should like to address is the compatibility or otherwise of the notion of rationality that stems from PTKL and *SA* as viewed from the perspective of the theoretician. If the discussion in the previous section has focused on the rationality of the

agents who are the object of modelling by the theoretician, this section focuses on the `rationality of the theoretician´ or, as Schumpeter (1984) denotes it, the `rationality of the observer´. We believe that the apparently irreconcilable notions of rationality that stem from PTKL and *SA* when looked at from the standpoint of the agents *consist, when approached from the standpoint of the theoretician, of the application of PTKL to two different problems*. First, we share Lagueux´s argument that `true´ rationality in Popper actually consists of the corrigibility of one´s beliefs (Lagueux, 2006, p. 202).[40] Second, and according to Popper, science represents a particular version of PTKL characterized by the application of the `critical method´:

`The difference between the amoeba and Einstein is that, although both make use of the method of trial and error or elimination, the amoeba dislikes erring while Einstein is intrigued by it: he consciously searches for his errors in the hope of learning by their discovery and elimination. The method of science is the critical method´ (Popper, 1972, p. 70).

Now, if science is characterized as the application of the `critical method´ to the object of knowledge, *we may rationalize SA as the specific application of the ´critical method´ to the social sciences*. That is, the theoretician of the social sciences proposes a reconstruction of the P-S in which actors find themselves by formulating a conjectural situational model where they make decisions on the basis of systematic deliberation. As such, the situational model proposed is only a conjecture according to which the former constitutes an `adequate´ oversimplification of the relevant P-S or:

`By conjectural analysis I mean a certain kind of tentative or conjectural explanation of some human action which appeals to the situation in which the agent finds himself… Admittedly, no creative action can ever be fully explained. Nevertheless, we can try, conjecturally, to give an idealized reconstruction of the problem situation in which the agent found himself, and to that extent make the action "understandable" (or "rationally understandable")…´ (Popper, 1972, p. 179).

The situational model constructed will then be subject to close scrutiny by the scientific community which will tell us the extent to which the different elements of the hypothetical P-S need to be modified. To the extent that the elements of the hypothetical P-S can be properly identified (i.e., it should consist of a set of observable elements such as physical and social constraints and assumptions about agents´ knowledge and information) both the situational model and the results of the empirical tests can be, in principle, openly criticized by the scientific community. That is, the significance of the situational model is that it is the `object´ against which criticism may be directed given the prior methodological decision to immunize *RP* from potential refutation. In turn, the role of *RP* (regardless of the specific version adopted) is to facilitate the implementation of the `critical method´ by helping scientists identify the `logic of the situation´ captured in the situational model. In this respect, it has been argued elsewhere that, in the context of Popper´s distinction among three ontological domains, the claim to *objectivity* in World 3 — which consists of knowledge or thought in an objective sense such as problems, theories, and arguments — stems from the notion that knowledge in World 3 `resides in *recorded* form outside the mind of any agent, even its originator, and is, in principle, accessible by any other agent in that recorded form´ (Oakley, 2002, p. 464; also Popper, 1972, pp. 108-9). As Sassower (2006, p. 104) has observed, Popper `saw rationality as the way to inter-subjectivity, because it is too much to expect objectivity´.

In the absence of an `animating principle´ such as *RP*, it would be very hard to logically connect the elements of the situational model in an understandable way.

What is crucial, however, is that the situational model will undergo successive changes and refinements (that will let it acquire greater accuracy in the explanation of social phenomena) in the aftermath of `rational´ criticism by the scientific community. Viewed from this standpoint, it does not make any difference whether the theoretician adopts *RPo* or *RPs* as long as the situational model is subject to close scrutiny by other members of the scientific community. Then, as the situational model undergoes further refinements or modifications new predictions and explanations may eventually emerge. It is clear then that, as long as social scientists modify their models in order to make them capable of providing increasingly accurate explanations of social phenomena, *their behaviour can be characterized as being `corrigible´ and, hence, as sticking to the type of rationality we have associated with PTKL*. We may thus conclude that, although there is some tension between PTKL and *SA* when their relation is approached from the standpoint of `the rationality of agents´ such tension does not arise when such relation is approached instead from the standpoint of the `rationality of the theoretician´.

5. Summary & conclusions

A number of commentators have noted that there are two different approaches to rationality in Popper´s philosophy: the approach stemming from his evolutionary theory of knowledge and learning (PTKL) and the approach embodied in so-called `Situational Analysis´ (*SA*) and associated to his famous `Rationality Principle´ (*RP*). According to the former, we `learn´ by subjecting our hypotheses to trial and discarding those ones which turn out to be wrong. In addition, all knowledge is conjectural and fallible. In this setting, science is the `highest´ form of knowledge acquisition and is characterized by the

subjecting of scientific theories to the most severe forms of criticism by members of the scientific community including, of course, empirical testing. The notion of `critical rationalism´ was coined to capture Popper´s thesis that the way to maximize the rate of expansion of knowledge is to subject theories to an optimal amount of criticism. The notion of human rationality that stems from PTKL consists of the *corrigibility* of our (wrong) beliefs. In other words, PTKL implies that being `rational´ consists of revising our beliefs when they turn out to be wrong. Therefore, individuals who fail to do so are `irrational´. *SA* represents Popper´s methodological proposal for the social sciences. In turn, *RP* is a methodological principle according to which *agents always act in a way that is adequate or appropriate to their problem-situation* (P-S). It follows that, in the context of *SA*, `rational´ behaviour consists of acting in a way that is appropriate to the `logic of the situation´ whereas `irrational´ behaviour will consist of doing otherwise. However, Popper and several of his commentators have made an important distinction between the `objectivist´ and the `subjectivist´ version of *SA*. According to them, in the former the theoretician seeks to reconstruct P-S `as it actually is´ whereas in the latter she reconstructs it `as it is seen by agents´. In this respect, we argued that the former is based on the (implicit) assumption that there is a systematic *discrepancy* between the theoretician´s and agents´ view of P-S and that, under the null hypothesis, this implies that agents´ view of P-S is assumed to be, at least partially, wrong. We also proposed a rationalization of the `objectivist´ version of *SA* according to which the latter is based on a methodological decision to assume that the mistakes made by agents when making decisions are *less* interesting for the purpose of understanding agents´ behaviour and, especially, for the generation of predictions than the (modelling) mistakes made by the theoretician.

The purpose of this essay was to study the compatibility of these two apparently irreconcilable approaches. We have made five different claims. Our first claim was that there is a certain tension between PTKL and *SA* when their relation is analysed from the standpoint of the `rationality of the agents´ whose behaviour is captured in the model albeit the tension disappears when the relation is analysed from the standpoint of the `rationality of the theoretician´. Our second claim was that the nature of the tension between PTKL and *SA* depends on whether the theoretician adopts the `objectivist´ or the `subjectivist´ version of *SA*. In particular, we argued that the tension between PTKL and the `subjectivist´ *SA* stems from the fact that, in the latter, it is implicitly assumed that agents´ view of P-S is partially wrong which implies, in turn, that agents do not `learn´ from their mistakes as PTKL posits, i.e., that they exhibit a tendency to repeat the same mistakes so that the latter become predictable. We argued that this feature of the `subjectivist´ version of *SA* creates a tension with PTKL when the main purpose of the theoretician is to generate predictions but does not generate any tension with PTKL when the purpose is to perform historical interpretation. This raises the difficult issue of the legitimacy of adopting the `subjectivist´ *SA* when the main purpose of the modelling exercise is to generate predictions the latter being understood as the derivation of the logical consequences of theory. It was not the purpose of this study to settle this issue and, hence, issuing a verdict on it will require further work. By contrast, we argued that the tension between PTKL and the `objectivist´ *SA* stems from the fact: (i) that if agents behave according to PTKL it is not necessarily the case that their decisions will be adequate or appropriate to the `logic of the situation´ insofar as the former *only* implies that agents tend to eliminate their mistakes and, hence, we have that in the wake of changes in the surrounding environment agents´ decisions may not be adequate to the `logic of the (new) situation´, and (ii) that the adoption of the `objectivist´ *SA* implies *de*

facto the *imposition* of the theoretician´s view of P-S upon agents´. Yet, we argued that it is unlikely that, if agents behave according to PTKL by subjecting their conjectures to trial and eliminating the ones that turn out to be wrong, their view of P-S will eventually converge to the theoretician´s. Our third claim built on the ideas of Hayek (1943) about the nature of the `facts´ of the social sciences and was that, in the way it is presented by Popper and some of his commentators, the `objectivist´ *SA* represents a *limit* or extreme case based on the presupposition that P-S is (fully) *independent* of agents´ beliefs. Our fourth claim was closely related to the previous one and consisted of the idea that, if Hayek´s ideas on the nature of the `facts´ of the social sciences are duly taken on board, it follows that the natural strategy for social scientists is to seek to reconstruct P-S *as agents´ see it* rather than to reconstruct it *as the scientist sees it*. Our fifth and last claim was that, unlike what Popper and his commentators suggest, the difference between the `objectivist´ and the `subjectivist´ version of *SA* is not that in the former the theoretician reconstructs P-S *as it actually is* whereas in the latter she does it as *agents see it* but, rather, that in the former she reconstructs P-S *as she sees it* herself whereas in the latter she does it *as she believes that agents actually see it.*

References

Alchian, A. A. (1950) Uncertainty, Evolution, and Economic Theory, *Journal of Political Economy*, 58(3), pp. 211-221.

Arrow, K. J. (1986) Rationality of self and others in an economic system, *Journal of Business*, 59(4.2): S385-S399.

Arrow, K. J. (2004) Is bounded rationality unboundedly rational? Some ruminations, in Augier, M. & March, J. G. (eds.) *Models of a Man: Essays in Memory of Herbert A. Simon*, Cambridge, MA: MIT Press, pp. 47–55.

Bartley, W. W. (1982) The Philosophy of Karl Popper. Part III. Rationality, Criticism, and Logic, *Philosophia* 11(1–2), pp. 121–221.

Becker, G. (1962) Irrational behavior and economic theory, *Journal of Political Economy*, 70(1), pp. 1-13.

Becker, G. S. (1976) *The economic approach to human behaviour*, Chicago (IL): University of Chicago Press.

Beinhocker, E. D. (2013) Reflexivity, complexity, and the nature of social science, *Journal of Economic Methodology*, 20(4), pp. 330-342.

Bentham (1907[1789]) *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press

Bicchieri, C. (1992) Two Kinds of Rationality, in de Marchi, N. (ed.) *Post-Popperian Methodology of Economics: Recovering Practice*, pp. 155-188, Boston: Kluwer Academic Publishers.

Blaug, M. (1975) Kuhn versus Lakatos, or Paradigms versus Research Programmes in the History of Economics, *History of Political Economy*, 7, pp. 399–419.

Blaug, M. (1985) Comment on D. Hands, 'Karl Popper and economic methodology: a new look', *Economics and Philosophy*, 1(2), pp. 286–8.

Boland, L. A. (2003*a*) Dealing with Popper in Economic Methodology, *Philosophy of the Social Sciences*, 33(4), pp. 479-98.

Boland, L. A. (2003*b*) *The Foundations of Economic Method: A Popperian Perspective*, 2nd edition, London: Routledge.

de Bruin, B. (2006) Popper's Conception of the Rationality Principle in the Social Sciences, in Ian Jarvie, David Miller & Karl Milford (eds.), *Karl Popper: A Centenary Assessment: Selected Papers from Karl Popper 2002: Volume III: Science*. Aldershot: Ashgate.

Bunge, M. (1996) The Seven Pillars of Popper´s Social Philosophy, *Philosophy of the Social Sciences*, 26(4), pp. 528-556.

Caldwell, B. J. (1985) Towards a broader conception of criticism, *History of Political Economy*, 18(4), pp. 675-81.

Caldwell, B. J. (1991) Clarifying Popper, *Journal of Economic Literature*, 29(1), pp. 1-33.

Caldwell, B. J. (2003) Popper and Hayek: Who Influenced Whom? *Cashiers D´Épistémologie*, 292, 2003-01.

Campbell, D. T. (1974) Evolutionary Epistemology, in Schilpp, P. A. (ed.) *The Philosophy of Karl Popper*, pp. 413-463, Vol. XIV, Book I, LaSalle (IL): Open Court.

Cosmides, L. & Tooby, J. (1994) Better than Rational: Evolutionary Psychology and the Invisible Hand, *American Economic Review*, 84(2), pp. 327-32.

De Bruin, B. (2006) Popper´s Conception of the Rationality Principle in the Social Sciences, in Jarvie, I., Minford, K. & Miller, D. (eds.) *Karl Popper: A Centenary Assessment*, Vol. III (Science), pp. 209-217, Ashgate.

Dow, S. C. (1995) The appeal of neoclassical economics: some insights from Keynes's epistemology, *Cambridge Journal of Economics*, 19(6), pp. 715-733.

Edgeworth (1881) *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. Augustus M. Kelley: New York.

Farmer, M. K. (1982) Rational Action in Economic and Social Theory: Some Misunderstandings, *Archives Europeennes de Sociologie*, 23(1), pp. 179-197.

Friedman, M. (1953) The methodology of positive economics, in *Essays in Positive Economics*, ch. 1, pp. 3-43, Chicago: Chicago University Press.

Gattei (2009) *Karl Popper's Philosophy of Science*, Routledge: New York.

Gibbard, A. & Varian, H. R. (1978) Economic Models, *Journal of Philosophy*, 75(11), pp. 664-677.

Goldthorpe, J. H. (1998) Rational Action for Sociology, *British Journal of Sociology*, 49(2), pp. 167-192.

Hamlin, A. P (1986) *Ethics, Economics and the State*, New York: St. Martin´s Press. Reprinted in Caldwell, B. (ed.), 1993, *The Philosophy and Methodology of Economics I*, pp. 304-362, Aldershot: Edward Elgar.

Hands, D. W. (1985) Karl Popper and Economic Methodology: A New Look, *Economics & Philosophy*, 1(1), pp. 83-99.

Hands, D. W. (1991) Popper, the rationality principle and economic explanation, in G. K. Shaw (ed.) *Economics, Culture and Education: Essays in Honour of Mark Blaug*, Cheltenham: Edward Elgar, pp. 108-119.

Hands, D. W. (1992) Falsification, Situational Analysis and Scientific Research Programs: The Popperian Tradition in Economic Methodology, in de Marchi, N. (ed.) *Post-Popperian Methodology of Economics: Recovering Practice*, pp. 19-54, Boston: Kluwer Academic Publishers.

Hayek, F. A. (1937) Economics and Knowledge, *Economica*, 4(13), pp. 33-54.

Hayek, F. A. (1942) Scientism and the Study of Society. Part I, *Economica*, 9(35), pp. 267-291.

Hayek, F. A. (1943) The Facts of the Social Sciences, *Ethics*, 54(1), pp. 1-13.

Hayek, F. A. (1967*a*) Kinds of Rationalism, in *Studies in Philosophy, Politics & Economics*, pp. 82-95, Chicago: University of Chicago Press.

Hayek, F. A. (1967*b*) Degrees of Explanation, in *Studies in Philosophy, Politics & Economics*, pp. 3-21, Chicago: University of Chicago Press.

Hedström, P., Swedberg, R. & Udéhn, L. (1998) Popper´s Situational Analysis and Contemporary Sociology, *Philosophy of the Social Sciences*, 28(3), pp. 339-364.

Hicks, J. R. (1956) *A Revision of Demand Theory*, Oxford: Clarendon Press.

Hobbes, T. (1985[1651]) *Leviathan*, London: Penguin Books.

Hume, D. (2006[1748]) *An enquiry concerning human understanding: a critical edition*, Oxford, Clarendon Press; New York: Oxford University Press.

Hume, D. (1875) *Essays Moral, Political & Literary*, vol. 2, ed. T.H. Green & T. H. Grose. London: Longmans Green & Co.

Hutchison, T. W. (1997) On the relation between philosophy and economics. Part II: To what kind of philosophical problems should economists address themselves? *Journal of Economic Methodology*, 4(1), pp. 127-151.

Jacobs, S. (1990) Popper, Weber and the Rationalist Approach to Social Explanation, *British Journal of Sociology*, 41(4), pp. 559-570.

Kerstenetzky, C. L. (2009) Plural situational logic: the rationa(lisabi)lity principle, *Cambrigde Journal of Economics*, 33(2), pp. 193-209.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. Cambridge: Macmillan.

Kirzner, I. (1976) *The Economic Point of View*, Kansas City: Sheed & Ward.

Klappholz, K. & Agassi, J. (1959) Methodological Prescriptions in Economics, *Economica*, 26(101), February, pp. 60-74.

Knight, F. H. (1971[1921]) *Risk, Uncertainty, and Profit*, Chicago: The University of Chicago Press.

Koertge, N. (1975) Popper's Metaphysical Research Program for the Human Sciences, *Inquiry*, 18(4), pp. 437-62.

Koertge, N. (1979) The Methodological Status of Popper´s Rationality Principle, *Theory and Decision*, 10(1/4), January, pp. 83-95.

Lagueux, M. (1993) Popper and the Rationality Principle, *Philosophy of the Social Sciences*, 23(4), pp. 468-480.

Lagueux, M. (2006) Popper and the Rationality Principle, in Jarvie, I., Minford, K. & Miller, D. (eds.) *Karl Popper: A Centenary Assessment*, Vol. III (Science), pp. 197-217, Ashgate.

Lagueux, M. (2010) *Rationality and Explanation in Economics*, Routledge.

Langlois, R. N. (1995) Rationality, institutions, and explanation, in R. N. Langlois (ed.) *Economics as a Process*, pp. 225-255, Cambridge: Cambridge University Press.

Langlois, R. N. (1998) Rule-following, Expertise, and Rationality: A New Behavioral Economics? in Dennis, K. (ed.) *Rationality in Economics: Alternative Perspectives*, pp. 55-78, Boston: Kluwer Academic.

Latsis, S. J. (1972) Situational Determinism in Economics, *British Journal for the Philosophy of Sciences*, 23(3), pp. 207-245.

Latsis, S. J. (1983) The Role and Status of the Rationality Principle in the Social Sciences, in Cohen, R. S. & Wartoksky, M. W. (eds.) *Epistemology, Methodology, and the Social Sciences*, pp. 123-152, London: Reidel Publishing Company.

Loasby, B. J. (1999) *Knowledge, Institutions and Evolution in Economics*, London: Routledge.

Loasby, B. J. (2003) Closed models and open systems, *Journal of Economic Methodology*, 10(3), pp. 285-306.

Lucas, R. E. (1986) Adaptive behaviour and economic theory, in R. M. Hogarth & M. W. Reder (eds.) *Rational Choice: The Contrast between Economics and Psychology*, pp. 217-242, Chicago and London: University of Chicago Press.

Matzner, E. & Jarvie, I. C. (1998) Introduction to the Special Issues on Situational Analysis, *Philosophy of the Social Sciences*, 28(3), pp. 333-338.

Mayr, E. (1988) *Toward a New Philosophy of Biology – Observations of an Evolutionist.* Cambridge, Mass., and London: Harvard University Press.

Menger, C. (1950[1871]) *Principles of Economics*, translated and edited by J. Dingwall & Bert H. Hoslitz. Glencoe, IL: Free Press.

Merton, R. K. (1948) The self-fulfilling prophecy, *Antioch Review*, 8(2), pp. 193-210.

Mill, J. S. (1967[1836]) On the Definition of Political Economy and the Method of Investigation Proper to It, in *Collected Works of John Stuart Mill*, vol. 4, pp. 309-39, Toronto: University of Toronto Press.

Minsky, H. P. (1975) *John Maynard Keynes*. New York: Columbia University Press.

Muth, J. F. (1961) Rational Expectations and the Theory of Price Movements, *Econometrica*, 29(3), pp. 315-335.

Nadeau, R. (1993) Confuting Popper on the Rationality Principle, *Philosophy of the Social Sciences*, 23(4), pp. 446-467.

Notturno, M. A. (1998) Truth, Rationality, and the Situation, *Philosophy of the Social Sciences*, 28(3), pp. 400-421.

Oakley, A. (1999) Economics and the origin of Popper´s situational analysis, *History of Economics Review*, 30, pp. 25-40.

Oakley, A. (2002) Popper´s Ontology of Situated Human Action, *Philosophy of the Social Sciences*, 32(4), pp. 455-486.

Pareto, V. (1917) *Traitè de sociologie generale*, Payot: Lausanne and Paris.

Parsons, T. (1937) *The Structure of Social Action*, McGraw Hill: New York.

Popper, K. R. (1966[1943*a*]) *The Open Society and its Enemies*, Vol. II, London: Poutledge & Kegan.

Popper, K. R. (1966[1943*b*]) *The Open Society and its Enemies*, Vol. I, London: Poutledge & Kegan.

Popper, K. R. (1944-45) The Poverty of Historicism I-III, *Economica*, 11, pp. 86-103, 119-137, 12: pp. 69-89.

Popper, K. R. (1963) *Conjectures and Refutations: The Growth of Scientific Knowledge*, London: Routledge and Kegan Paul.

Popper, K. R. (1966) *Of Clouds and Clocks: An Approach to the Problem of Rationality of and the Freedom of Man*, St. Louis: Washington University Press.

Popper, K. R. (1967) La rationalité et le statut du principe de rationalité, in E. M. Classen (ed.), *Les Fondements Philosophiques des Systèmes Economiques*, pp. 142-50, Paris: Payot.

Popper, K. R. (1972) *Objective Knowledge: An Evolutionary Approach*, Oxford: Clarendon Press.

Popper, K. R. (2002[1976*a*]) *Unended Quest*, Routledge: New York.

Popper, K. R. (1976*b*) The Logic of the Social Sciences, in T. W. Adorno et al., *The Positivist Dispute in German Sociology*, pp. 87-104, London: Heinemann.

Popper, K. R. (1985) The Rationality Principle, in D. Miller (ed.) *Popper Selections*, pp. 357-365, Princeton: Princeton University Press.

Popper, K. R. (1990) *A World of Propensities*, Bristol (UK): Thoemmes.

Popper, K. R. (1994) *The Myth of the Framework: In defence of science and rationality*, London: Routledge.

Sassower, R. (2006) *Popper´s Legacy: Rethinking Politics, Economics and Science*, Stocksfield (UK): Acumen.

Savage, L. J. (1954) *The Foundations of Statistics*, New York: Wiley.

Savage, L. J. (1962) *The Foundations of Statistical Inference*, London: Methuen & Co. Ltd.

Schumpeter, J. A. (1984) The Meaning of Rationality in the Social Sciences, *Journal of Institutional and Theoretical Economics*, 140(4), pp. 577-593.

Schutz, A. (1972) *The Phenomenology of the Social World*, London: Heinnemann.

Simon, H. A. (1976) From Substantive to Procedural Rationality, in Spiro J. Latsis (ed.) *Method and Appraisal in Economics*, pp. 129-148, Cambrigde: Cambridge University Press.

Smith, A. (1980[1795]) The principles which lead and direct philosophical enquiries: illustrated by the history of astronomy, *Essays on Philosophical Subjects*, ed. W.P.D. Wightman, Oxford: Oxford University Press.

Twomey, P. (1998) Reviving Veblenian Psychology, *Cambridge Journal of Economics*, 22(4), pp. 433-48.

Tversky, A. & Kahneman, D. (1986) Rational choices and the Framing of Decisions, in R. M. Hogarth & M. W. Reder (eds.) *Rational Choice: The Contrast between Economics and Psychology*, pp. 67-94, Chicago: University of Chicago Press.

Vanberg, V. J. (2002) Rational Choice vs. Program-based Behaviour: Alternative Theoretical Approaches and their Relevance for the Study of Institutions, *Rationality and Society*, 14(7), pp. 7-54.

Volz, K. H. & Gigerenzer, G. (2012) Cognitive processes in decisions under risk are not the same as in decisions under uncertainty, *Frontiers in Neuroscience*, 6(1), pp. 1-6.

von Mises, L. (1944) The Treatment of "Irrationality" in the Social Sciences, *Philosophy and Phenomenology Research*, 4(4), pp. 527-546.

von Neumann, J. & Morgenstern, O. (1947) *Theory of Games and Economic Behaviour*, Princeton (NJ): Princeton University Press.

Vromen, J. J. (1995) *Economic Evolution: An Enquiry into the Foundations of New Institutional Economics*, London: Routledge.

Weber, M. (1904-5) *The Protestant Ethic and the Spirit of Capitalism*, London: Allen & Unwin.

Weber, M. (1949) *The Methodology of the Social Sciences*, New York: Free Press.

Wible, J. R. (1984/85) An epistemic critique of rational expectations and the neoclassical macroeconomics research program, *Journal of Post-Keynesian Economics*, 7(2), pp. 269-81.

Zouboulakis, M. S. (2014) *The Varieties of Economic Rationality: From Adam Smith to contemporary behavioural and evolutionary economics*, Routledge.

Notes

[1] General criticisms of Popper´s methodological proposal for the social sciences can be found in Latsis (1983), Hands, (1985), Bunge (1996), Hutchison (1997), Oakley (1999, 2002), and Vanberg (2002).

[2] Recently, another version of *RP* has been proposed in Lagueux (2006, pp. 201-202) who suggests that, given the fact that refinements in model-construction in the social sciences imply that theoretical models exhibit more detailed descriptions of the situation, *RP* may also be enunciated as implying that `the agent will agree with what is clearly presented by the model itself as the appropriate thing to do´. Be that as it may, for the purpose of this study we focus hereafter on Popper´s version of *RP*.

[3] These ideas were expounded in his famous *Logic of Scientific Discovery*. That said, and as far as the social sciences are concerned, there is some consensus on the notion that it is not 'falsificationism' *per se* but 'critical rationalism' — of which falsification is only one possibility — that is the true message of Popper´s philosophy (Caldwell, 1991; Notturno, 1998; Boland, 2003*a*). According to Hands (1991, p. 114), who nevertheless remains sceptical of this interpretation, `critical rationalism´ is an interpretation of Popperian philosophy due primarily to Bartley (1982).

[4] In his analysis of the role and status of *RP* Latsis (1983) focuses on Popper´s analysis of the `mind-body problem´, that is, the analysis of the manner in which mental states affect behaviour as discussed in his

paper `Of Clouds and Clocks´ (Popper, 1966). According to Latsis (1983, p. 139), *RP* represents Popper´s compromise solution to this problem whereby it is suggested that `our mental states control some of our behaviour and that this control is "of a plastic kind"´

[5] Interestingly, Hayek (1967*a*, p. 86) associates the notion of rationality in Classical economics to the ability to `learn from experience´ which is very close to PTKL.

[6] Incidentally, Loasby (2003, pp. 296-7) argues that Smith (1980[1795]) replaced Hume´s argument by a proto-Popperian theory of the growth of knowledge according to which all observed patterns that exhibit a sufficient degree of similarity are treated as (fallible) imaginative conjectures.

[7] Popper adopted Tarski´s theory that truth is correspondence with the facts or with reality (Popper, 1972, p. 44).

[8] Popper´s notion of `subjective knowledge´ also bears a strong resemblance to the notion of 'reasoning instincts' used in evolutionary psychology (Cosmides & Tooby, 1994, p. 330).

[9] According to Gattei (2009, p. 58), Popper´s World 3 bears a strong similarity to Plato´s theory of Ideas, and to Hegel´s theory of the Objective Spirit although he thinks it is `closer to Bolzano´s theory of a universe of statements in themselves and truth in themselves, or to Frege´s universe of objective contents of thought´.

[10] As he aptly notes, a tautology, though obviously true, has zero *truth content* and zero verisimilitude.

[11] Of course, this applies to any premise aimed at providing a 'closure' for a model. As Loasby (1999, p. 14) reminds us, 'all closures are in some degree false. There can be no self-sufficient Cartesian scheme for deducing justified true knowledge from some original certainty'.

[12] This touches upon the issue of the incompatibility of Popper´s *RP* and his falsificationist methodology. For instance, Caldwell (1991, p. 13) argues that 'Popper´s rationality principle represents an immunizing stratagem that is elevated to the status of an inviolable methodological principle'. In an attempt to reconcile both principles, Koertge (1979, p. 93) interprets *RP* as the Lakatosian hard-core of Popper´s research program in the social sciences whereas the positive heuristic is 'his metaphysical theory of man as an evolving rational problem-solving animal'.

[13] Notwithstanding the intellectual authority of Popper, economists like Hayek (1967, p. 29) or Hutchison (1977, p. 43) question the applicability of Popper's falsificationanist methodology to economics. Indeed, the applicability of strict falsificationism to the social sciences appears to be problematic even to Popper himself as noted in Hands (1985, p. 96).

[14] However, we will argue below that a clear antecedent of this distinction is in Hayek (1948[1937]).

[15] By contrast, we will argue below that the `subjectivist´ version of *RP* is a legitimate and potentially fruitful one in the social sciences.

[16] Yet, Zouboulakis (2014, p. 87) argues that it is clear that Popper has an `objectivist´ version of *RP* in mind.

[17] Prior to this clarification, Popper (1972) had already recognized the existence of two versions of *SA*:

> `There are many cases in which we can reconstruct, *objectively* (even though conjecturally), (a) the *situation as it was* and (b) a very different *situation as it appeared to the agent*, or as it was *understood, or interpreted* by the agent. It is interesting that this can be dome even in the history of science´ (Popper, 1972, p. 179, footnote 27).

[18] However, it has been argued that, although the notion of `single-exit´ modelling and *SA* originates with Weber, for him this and related concepts were tools of *historical* analysis, not of theory (Langlois, 1995, p. 230; 1998, p. 69). Interestingly, a philosophical foundation for Weber´s notion of `ideal type´ is in the school of phenomenology and, particularly, in Schutz´s concept of `second-order typifications´ (Schutz, 1972).

[19] The interrelation between the ideas of Hayek and Popper is discussed in Oakley (1999) and Caldwell (2003). For instance, Popper (1966[1943a])) appears to have been inspired by Hayek (1942) when expressing his crucial idea that both our institutions and traditions are largely the `*indirect, the unintended and often the unwanted by-products*´ of conscious and intentional human actions and, therefore, that `only a minority of social institutions are consciously designed, while the vast majority have just "grown", as the undesigned results of human actions´ (Popper, (1966[1943a]), p. 93). Likewise, Hayek had already argued that social studies deal `not with the relations between things, but with the relations between men and things or the relations between man and man. They are concerned with man´s actions and their aim is to explain the unintended or undesigned results of the actions of many men´ (Hayek, 1942, p. 276).

[20] In this respect, let us mention that Menger (1950[1871]), p. 148) was one of the first social scientists to incorporate error within his model and to argue that *all* knowledge (both of the actors and of the theorist) is bound to be prone to errors.

[21] As von Mises explains:

`Every human action aims at the substitution of more satisfactory conditions for less satisfactory. Man acts because he feels uneasy and believes that he has the power to relieve to some extent his uneasiness by influencing the course of events. A man perfectly content with the state of his affairs would not have any incentive to change things; he would have neither wishes nor desires, he would not act because he would be perfectly happy... Strictly speaking, only the increase in satisfaction (decrease of uneasiness) should be called *end*, and accordingly all states which bring about such an increase *means´* (von Mises, 1944, p. 532).

[22] The notion of `instrumental rationality´ goes back as far as Max Weber´s sociological histories of world religions culminating with his classic study of modern European Christianity (Weber, 1904-5).

[23] For instance, in standard consumer theory, 'means-rationality' consists of the axioms of completeness, independence, reflexivity, and transitivity of preferences (Dow, 1995, p. 724). In the case of SEU theory developed by von Neumman & Morgenstern (1947) and, especially, Savage (1954), which represents the canonical theory of choice under conditions of uncertainty, 'means-rationality' is satisfied if the following four assumptions or axioms are fulfilled: cancellation, transitivity, dominance, and invariance. To this, we may add the more technical assumptions of comparability and continuity.

[24] In fact, Savage (1962, p. 14) refers to the notion of objectivity in knowledge as not `valid, fruitful or practical´.

[25] As noted in Beinhocker (2013), in the context of the social sciences, a prediction (unlike a forecast) amounts barely to the *deductive logical consequences* of a theory. It should be noted, however, that the predictions to be derived from the models in the social sciences differ from the predictions generated by the theories of the natural sciences. Building on Hayek´s distinction between `explanation in principle´ and `explanation in detail´ (Hayek, 1967*b*, p. 20), Popper (1994, p. 163) distinguishes between explaining or predicting *singular events* from problems of explaining or predicting a *kind* or *type* of event. According to Popper, the former can be solved *without constructing a model* **—** no more than certain universal laws and the relevant initial conditions are needed **—** whereas the latter is most easily solved *by means of constructing a model* (*op. cit.*, p. 164). Further, he argues (Popper, *op. cit.*, p. 165), that a model consists of `certain elements placed in a typical relationship to each other, plus certain universal laws of interaction — the "animating" laws´. Unlike theories, models try to capture the *typical* aspects of P-S in order to make statements about a *type* of event and, hence, they represent something akin to *typical initial conditions* (*op. cit.*, p. 164). In turn, a statement about a *typical* event can be either an explanation of *why* that typical event occurred in the past, or else, a prediction, that is, a logical consequence of the theory.

[26] However, the generation of predictions does not seem to be, according to Popper, the main aim of the social sciences. For instance, according to Notturno (1998, p. 412, emphasis added), the `problem of situational analysis in the theoretical and historical social sciences, in Popper´s view, is not to construct models that predict or prophesize the future; it is to construct models that help us to *explain and understand the past´*. In this respect, Popper seems to view *RPs* as being particularly suited to the task of historical explanation:

`The historian´s task is, therefore, so to reconstruct the problem situation as it appeared to the agent, that the actions of the agent become *adequate* to the situation… Our conjectural reconstruction of the

situation may be a real historical discovery. It may explain any aspect of history so far unexplained´ (Popper, 1972, p. 189).

To be sure, Popper (*op. cit*., p. 166, emphasis in original) explains that `the fundamental problem of both the theoretical and the historical social sciences is *to explain and understand events in terms of human actions and* [typical] *social situations´* and, hence, does not mention the generation of *predictions* as the aim of the social sciences.

[27] It is clear that Popper admired neoclassical economics. For instance, he writes that:

`The social sciences never had for me the same attraction as the theoretical natural sciences. In fact, the only theoretical social science which appealed to me was economics´ (Popper, (2002[1976a]), p. 121).

Notwithstanding it, Blaug (1985, p. 287) argues that `Popper knew little about social sciences and less about economics´.

[28] Similarly, Hutchison (1997) does not distinguish between *RPo* and *RPs* yet his critical comments on *RP* seem to have *RPo* as his target. In particular, he refers to a `fortified´ version of *RP* that is pervasive in mainstream economics which includes the (very unrealistic) `full-knowledge´ assumption without which, he argues, *RP* is `almost empty´. Notwithstanding it, he argues that in its 'fortified' version, *RP* `seems a good or fair approximation in quite a range of cases, but no approximation at all in another important range of cases´ (*op. cit*., p. 139).

[29] He also refers to the `relative attractiveness´ for an individual of different actions (Savage, 1951, p. 61).

[30] Savage (1954) discusses the scope of his theory of rational decision-making and his notion of `small world´ in chapter 2:

`The argument might be raised that the formal description of decision that has thus been erected seems inadequate because a person may not know the consequences of the acts open to him in each state of the world. He might be so ignorant, for example, as not to be sure whether one rotten egg will spoil a six-egg omelette. But in that case noting could be simpler than to admit that there are four states in the world corresponding to the two states of the egg and the two conceivable answers to the culinary question whether one bad egg will spoil a six-egg omelette…

What in the ordinary way of thinking might be regarded as a chain of decisions, one leading to the other in time, is in the formal description proposed here regarded as a single decision. To put it a little differently, it is proposed that the choice of a policy or plan be regarded as a single decision…

The point of view under discussion may be symbolized by the proverb, "Look before you leap," and the one to which it is opposed by the proverb, "You can cross that bridge when you come to it"…

Carried to its logical extreme, the "Look before you leap" principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in the light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous, not — as some might think — because there might be later cause for regret, if things did not turn out as had been anticipated, but because the task implied in making such a decision is not even remotely resembled by human possibility. It is even utterly beyond our power to plan a picnic or to play a game of chess in accordance with the principle, even when the world of states and the set of available acts to be envisaged are artificially reduced to the narrowest reasonable limits…

Though the "look before you leap" principle is preposterous if carried to extremes, I would none the less argue that it is the proper subject of our further discussion, because to cross one´s bridges when one come to them means to attack relatively simple problems of decision by *artificially confining attention to so small a world* that the "Look before you leap" principle can be applied there…´ (Savage, 1954, pp. 15-16, emphasis added).

However, according to Volz & Gigerenzer (2012, p. 1), most of the time we make decisions under Knightian uncertainty, while situations of risk are relatively rare and found mostly in gambling. They also note that Savage (1954, p. 16), makes it clear that applying Bayesian theory to decisions in uncertain or Knightian worlds would not make sense because there is no way to know all the alternatives and its likely consequences. Likewise, Arrow (2004, p. 54) argues that in uncertain, ill-specified worlds, maximization of expected utility 'has no meaning at all'.

[31] This can be ascribed ultimately to the fact that, as Savage (1954, p. 20) recognizes, `logic itself admits an empirical [or descriptive] as well as a normative interpretation´.

[32] For instance, Farmer (1982, p. 179) argues that the `rationality assumption´ is in the `hard core´ of the (Lakatosian) economist´s research programme. Matzner & Jarvie (1998) even suggest that Popper´s *SA* represents a *soft* version of `economic imperialism´ yet to be distinguished from a *strong* version they associate with the work of Gary Becker. In any case, they recognize that it was Popper — and not Gary Becker — who first formulated a programme for extending the logic of economics to the noneconomic social sciences (*op. cit*., p. 336).

[33] An alternative rationale to agents´ learning process for the adoption of the optimization assumption in mainstream economics is the 'natural' selection argument. This argument was originally proposed by Alchian (1950) in the context of economic competition and since then it has been advocated by a number of neoclassical theorists. Alchian (*op. cit*.) views the economic system as an *adoptive* mechanism which cleverly chooses among actions generated by the adaptive pursuit of `profits´. His argument starts with the premise that the realization of profits is the criterion according to which successful firms are selected. However, he admits that this process may be *independent* of the nature of the decision-making processes of economic agents. In particular, he recognizes that economic success does not require proper motivation but may instead be the outcome of fortuitous circumstances. Further, to the extent that profits accrue only to those firms who are better than their competitors, what matters for survival is one's competitiveness *relative* to the others rather than their absolute proximity to an optimum. Nevertheless, he recognizes that conscious or purposeful behaviour (in addition to sheer chance) also plays a role in the selection of firms by the market. Two such examples of purposeful behaviour are: (i) imitation of strategies previously adopted by successful firms, and (ii) the adoption of trial and error strategies aimed at improving one's adaptation to the environment. However, he makes it clear that the former should *not* be understood as mechanisms through which adequate or `rational´ actions can be selected thereby allowing firms to converge to an optimum in the form of profit maximization. This is because the latter would require the fulfilment of the two following convergence conditions (Alchian, 1950, p. 219): (i) that every single trial must be classifiable either as a success or as a failure, and (ii) the continual rising toward some *optimum optimorum* without the occurrence of intervening descents. In this respect, he writes:

'These convergence conditions do not apply to a changing environment, for there can be no observable comparison of the result of an action with any other… As a consequence, the measure of goodness of actions in anything except a tolerable-intolerable sense is lost, and the possibility of an individual´s converging to the optimum activity via a trial-and-error process disappears. Trial and error becomes survival or death. It cannot serve as a basis of the *individual´s* method of convergence to a "maximum" or optimum position.' (*op. cit*.).

The upshot of Alchian´s discussion is that successful adaptation within a stable environment may give *ex-post* the appearance of rational or optimizing behaviour at the individual level *even though no ex-ante rational calculation actually occurred*. Further, and as Loasby (1999, pp. 20-21; see also Vromen, 1995, pp. 32-33) insists, the economic `survival´ argument can only allow its advocates to claim that surviving firms will have achieved results which are, on average, *closer* to the maximisation of profits than those firms that did not survive and, hence, it does not allow them to use it as a justification for the occurrence of optimization at the *individual* level as Friedman (1953, p. 22), for instance, does. Although with significant qualifications, developments of Alchian´s argument can be found in Friedman (1953) and Becker (1962) and critical assessments of the `survival´ argument are in Vromen (1995), Loasby (1999) and Lagueux (2010).

[34] By contrast, Arrow (1986, p. S385) criticizes the view that the optimizing assumption can be justified as being the result of a process of learning and adaptation.

[35] Wible (1984-85, p. 271) characterizes the approach to decision-making and expectations formation embedded in neoclassical economics as one of *instantaneous rational assessment* and hypothesizes that its origin is the emphasis in the 'logic of justification' of knowledge rather in the 'process of discovery' of knowledge made by proponents of the logical positivist school of philosophy.

[36] Although Hayek (1967*a*, p. 89) does not mention the term SR explicitly, he seems to have this notion in mind when he characterises the approach to rationality inherent in the philosophical tradition he denotes as `constructivist rationalism´ and which is no other that the Cartesian tradition which is directly related to Descartes and goes back at least as far as Plato in Ancient Greece and whose main representatives are, according to him, Hobbes, Rousseau, Voltaire, Hegel and Marx. In particular, he describes the approach to rationality embodied in this approach as follows:

`In moral philosophy the constructivist rationalism tends to disdain any reliance on abstract mechanical rules and to regard as *truly rational* only behaviour such as is based on decisions which judge each particular situation "on its merits", and chooses between alternatives in concrete evaluation of the known consequences of the various possibilities´ (*op. cit*., emphasis added).

[37] Note the similarity of this definition with Koertge´s (1975) reformulation of *RP* provided above.

[38] PR is coupled to its sister notion of 'bounded' rationality (BR). According to Simon (1979), economic agents´ knowledge is subject to three different types of constraints: (i) limited ability to process, analyse, and store information, (ii) uncertainty, and (iii) the presence of social institutions. BR stems from the fact that the existence of these constraints prevents economic agents from 'optimizing'. Simon coined the term 'satisficing' to denote a decision-making rule that attempts to meet an acceptability (minimum) threshold. In contrast to 'satisficing' behaviour, the purpose of optimal decision-making is to find the best option available.

[39] This idea is discussed in detail also in Hayek (1942).

[40] For instance, in *The Open Society and its Enemies*, Popper (1966[1943a], p. 97) states that `the method of applying a situational logic to the social sciences is not based on any psychological assumption concerning the rationality (or otherwise) of "human nature"´. In this respect, Oakley (1999, p. 35) notes that Popper stresses in his Harvard lecture (Popper, 1985, 1994) that, in the context of *SA*, `rationality had no ontological significance and was not intended as a theory of human action´.

II. Keynes´s Approach to Macroeconomic Modelling: a Popperian Reconstruction

`Orthodoxy is the death of knowledge, since the growth of knowledge depends entirely on the existence of disagreement' (K. Popper, *The Myth of the Framework: In defence of science and rationality*, 1994, p. 34).

1. Introduction

As we showed in the previous essay, Hume was sceptical about the rationality of human behaviour since he believed that individual behaviour was grounded on `custom and habit´, that is, on the assumption that the `future resembles the past´ in spite of the inexistence of an `inductive logic´. This paradoxical situation was denoted by Hume as the `problem of induction´. Hume´s philosophical analysis remained unchallenged for more than two hundred years as Keynes (1920, pp. 303-04) recognized. In the previous essay we discussed the `solution´ to the problem of induction propounded by Popper which we denoted as Popper´s evolutionary theory of knowledge and learning (PTKL). However, the focus of our discussion was not so much on PTKL itself as on the tension with Popper´s methodological proposal for the social sciences referred to as `Situational Analysis´ (SA). Chronologically, there is a prior-to-Popper attempt to provide a solution to Hume´s `problem of induction´: Keynes´s (failed) attempt to construct an objective epistemological theory of probability in his *Treatise on Probability* (Keynes, 1920). As we will see, Keynes´s aim was not to only to answer Hume but to answer Cambridge philosopher G. E. Moore (1993[1903]) who argued that the highest expected `good´ would result if individual behaviour consisted of following rules insofar as the latter

represent the accumulated knowledge in our society. Keynes´s proposal to solve the `problem of induction´ in his *Treatise* consists of replacing the notion of `certain´ knowledge, which Hume showed to be unattainable, by the notion of `probabilistic´ knowledge. Keynes´s main argument was that, though we cannot attain the former, we can nevertheless attain the latter if only we adopt inductive procedures. The main innovation in the *Treatise on Probability* are logical probability relations. As noted in Andrews (1999), the latter look like Platonic Universals in the sense that they exist in a `logical space´, are independent of individual opinions and can only be discovered via intuition. In the *Treatise*, Keynes proposes an objective epistemological theory of probability. As we will see below, such proposal is grounded upon an inductive theory of knowledge and learning. However, the notion of logical probability relations was criticized by Ramsey (1978[1931]) in his *Truth & Probability* and Keynes ended up yielding to Ramsey´s criticisms in `My Early Beliefs´ published in 1938 (Keynes, 1973*a*, p. 445).

Next, the textual evidence suggests that by the second half of the 1930s Keynes had abandoned most of the ideas he had advocated in the *Treatise* and adopted instead a pragmatic approach to human behaviour which, as we shall show below, is remarkably similar to Hume´s views. Indeed, Keynes´s attempt to solve the `problem of induction´ in his *Treatise* was a failure and this ultimately forced him to return to the point where Hume had left off about two hundred years earlier. The return to Hume is transcendental for the methodology of economics since, as some commentators recognize, Keynes´s later economic writings are based on a theory of human nature that is similar to Hume´s (Davis, 1994; Andrews, 1999). Therefore, we take Hume´s theory of human nature as a key feature of Keynes´s later economic writings, including the *General Theory* (Keynes, 1936) and his paper in the *Quarterly Journal of Economics* (Keynes, 1937). This view

consists of the notion that people make decisions on the basis of inductive procedures which amount to adopting social rules and conventions which are widely believed to have worked well in the past. The rules and conventions Keynes refers to in his later economic writings are, for some purposes, equivalent to Hume´s old idea that individual behaviour is grounded on `custom and habit´. In particular, the chief convention of all, according to Keynes, is the assumption that `the existing state of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change´ (Keynes, 1936, p. 152).

The next issue we address in the essay is the compatibility of Keynesian macro-theory with PTKL. This is an issue which, to the best of our knowledge, has not been addressed in the literature on Keynesian economics. We make several claims. First, the replacement of the assumption that economic agents assume, unless they have specific reasons to think otherwise, that the `future will resemble the past´ can be conceptualized in a way that makes Keynesian macro-theory congenial to PTKL. More specifically, we propound the notion of `hegemonic´ conjectures (HC) to denote the social conventions which are central to Keynes´s macroeconomic analysis in his later economic writings. If we interpret HC as conjectures which are held *provisionally* by economic agents as long as the former believe that they provide valid guidelines upon which to ground decisions, then, it will be argued, the notion of HC provides a *link* between PTKL and Keynesian theory. As our argument goes, HC will be kept as long as they yield good results and will be `scrapped´ if agents believe that this is no longer the case. This implies that HC exhibit an *evolutionary* dimension in that they are retained by economic agents as long as the former appear to aid decision-making in a context of genuine uncertainty yet they are replaced by other (new) HC if economic agents no longer believe they provide a valid basis for decision-making. Now, the replacement of `old´ HC will take place in the aftermath of the occurrence of a substantial discrepancy between expected and realized

outcomes. This, in turn, will set off the search for new HC. We will denote the process of search for `new´ HC as a `confidence´ crisis in the sense that the `old´ HC have been abandoned but the `new´ HC have not yet emerged. Thus, we view `confidence´ crisis in a way that is, arguably, reminiscent of T. Kuhn´s theory of scientific revolutions (Kuhn, 1962). Lastly, we argue there is a parallelism between the notion of `confidence´ crisis and the explanation of economic downturns in Keynes´s macro-theory. In the latter, a downturn is heralded by the observation that there is a very large discrepancy between expected and realized outcomes. In turn, this brings about a fall in `confidence´ and an increase in liquidity preference. In short, we argue that the notion of HC provides a link between Keynes´s macro-theory and Popper´s theory of knowledge and learning in that, in both cases, the expansion of knowledge (conventional or otherwise) over time is the result of a process of trial and error-elimination.

Next, the convention that `the existing state of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change´ is at the very core of Keynesian macro-theory and is, in turn, responsible for two criticisms Keynesian theory has frequently been subject to[1]: (i) the alleged `irrational´ behaviour of the agents whose behaviour Keynesian models seek to capture, and (ii) the *nihilism* of Keynesian theory. To this, we may add another potential criticism that stems from the fact that, on the one hand, Keynes — in his criticism of the pioneering work of Tinbergen — argues that the use of statistical inference methods is valid in economics only if we can show that the probability distributions of economic variables exhibit a sufficient degree of stability over time whereas, on the other hand, his macro-theory is based upon the assumption that economic agents *believe*, unless they have good reasons to believe otherwise, that `the existing state of affairs will continue indefinitely´.

Now, we will argue below that the three criticisms referred to above are a direct implication of the adoption by mainstream theoreticians of the `objectivist´ version of Popper´s `Rationality Principle´ (*RPo*) we discussed in the previous chapter. First, we argued in that chapter that the adoption of *RPo* implies that the theoretician reconstructs the problem-situation (P-S) `as she believes it is´. In the case of mainstream theory, the former implies that agents are assumed to exhibit `ends-rationality´, `means-rationality´ and, `beliefs-rationality´. Now, the assumption of `beliefs-rationality´ implies, in turn, that economic agents´ view of P-S is assumed to be similar to the theoretician´s. In such circumstances, the theoretician can `close´ the model in a way that a number of testable predictions can be extracted. Now, we argue below that mainstream economists assess Keynes´s theory under the (often implicit) assumption that there must be a `coincidence´ between the agents´ and the theoreticians´ view of P-S. Thus, if the theoretician does *not* believe that `the future will resemble the past´ — as Keynes does — it follows that the construction of theoretical models where economic agents *believe* that `the existing state of affairs will continue indefinitely, except in so far as they have specific reasons to expect a change´ is, from the viewpoint of mainstream economists, flawed. Second, the accusation of `nihilism´ is related to the fact that the possibility of generating clear testable predictions in Keynes´s theory is, according to some mainstream economists, precluded by the fact that economic agents´ expectations are exogenous and may change capriciously. However, as we will argue below, the latter also stems from the fact that, unlike in neoclassical economics, in Keynes´s macro-theory the theoretician does not `impose´ her view of P-S upon the agents. Lastly, the accusation of positing theoretical models where the behaviour of economic agents is `irrational´ or *ad hoc* is also a direct consequence of the fact that, in Keynes´s macro-theory, agents exhibit `ends-rationality´ and `means-rationality´ but *not* `beliefs-rationality. In other words, Keynesian economic

agents are not optimizers and, hence, from the standpoint of neoclassical theory, this is tantamount to `irrational´ behaviour (Becker, 1962). Nevertheless, we will argue that the failure to optimize by Keynesian economic agents ultimately stems from the fact that, whereas neoclassical theory presumes that there exists an `inductive logic´, Keynes´s macro-theory presumes the opposite.

Now, we argued in the previous chapter that the adoption of the `subjectivist´ version of Popper´s Rationality Principle (*RPs*) implies that the theoretician reconstructs P-S `as she believes that agents believe it is´. We will argue below that, if Keynesian macro-theory is interpreted as an instance of *RPs*, the above-mentioned criticisms are unwarranted. First, even if the theoretician does *not* really believe that `the existing state of affairs will continue indefinitely´, she may nevertheless assume that economic agents believe so, or else, that they believe this is the best assumption they can make owing to the presence of uncertainty. Indeed, people make decisions every day in spite of the fact that, in many circumstances, they are uncertain of their future consequences. It follows that the interpretation of Keynesian macro-theory we propound based upon its implicit adoption of *RPs* implies, arguably, that there is no inherent contradiction in assuming that economic agents believe that `the existing state of affairs will continue indefinitely, except in so far as they have specific reasons for expecting a change´ even though the theoretician does not believe so. Second, if it is assumed that economic agents believe for practical purposes that the `future will resemble the past´, then Keynesian theory cannot be accused of portraying individual behaviour as `irrational´ since the former is *coherent* with agents´ belief in the presence of a high degree of stability in economic affairs. Last, Keynesian macro-theory cannot be accused of `nihilism´ either since the assumption that economic agents believe that `the existing state of affairs will continue indefinitely, except in so far as they have specific reasons to expect a change´ enables the theoretician

to construct macro-models that yield testable *predictions* (as opposed to forecasts) by which we mean the `logical implications´ of the theory.

The content of the essay is as follows. The following section presents Keynes´s approach to probability and expectations as presented in his *Treatise on Probability*. In section 3, we expound Ramsey´s critique on the notion of probability Keynes advocates in the *Treatise* as well as other relevant aspects of Ramsey´s thought as presented in his essay *Truth & Probability*. In section 4, we present the approach to expectations Keynes adopts in his later economic writings in the wake of his capitulation to Ramsey´s views as well as his views on the nature of economics and its object of study. We propose the notion of HC to illustrate the thesis that the theory of knowledge and learning implicit in Keynes´s macro-theory is compatible with PTKL. In section 5, we argue that some of the criticisms of Keynesian macro-theory by mainstream economists can be rationalized as stemming from the fact that the latter implicitly adopt the `objectivist´ version of *RP* whereas, according to us, Keynesian macro-theory is an instance of the `subjectivist´ version of *RP*. Further, we argue that, if reinterpreted in this fashion, Keynesian theory can answer most, if not all, the charges levelled at it. Finally, section 6 summarizes and concludes.

2. The Treatise on Probability

This section is devoted mainly to presenting Keynes´s philosophical work in his *Treatise on Probability* (Keynes, 1920). As we will see, this work stems partly from the attempt by Keynes to find a solution to the `problem of induction´ as posited by Hume two hundred years earlier and, partly, from an attempt to answer Cambridge philosopher G. E. Moore. There is an ongoing debate about the relevance of Keynes´s philosophical work for understanding his later economic writings. However, although we will refer in

passing to this debate at different points in our discussion, one reason for reviewing this episode of Keynes´s intellectual journey is that there are three key elements of Keynes´s later economic writings which are already present in his *Treatise on Probability*: (i) the possibility that some probabilities are unmeasurable, (ii) the notion of the `weight of the argument´, and (iii) the rejection of the frequency approach to probability in the social sciences. Another reason for reviewing the *Treatise* is that, in the aftermath of Keynes´s capitulation to Ramsey´s (successful) criticism of the notion of objective probability relations, Keynes implicitly adopted in his later economic writings a theory of human nature similar to Hume´s. In particular, Hume´s conclusion that human behaviour is not `rational´ insofar as it is grounded on `custom and habit´ could be applied, in principle, to Keynes´s theory of human behaviour as based on conventional knowledge. However, we will argue in subsequent sections that this is not the case since Keynes´s later macro-theory can be rationalized in the light of both PTKL and *RPs*.

2.1. The ethical roots of Keynes´s *Treatise on Probability*

Bateman (1988) argues that Keynes´s *Treatise on Probability* is, at least partly, an answer to Moore´s *Principia Ethica*. Moore (1993[1903]) believed that our actions should be evaluated solely in terms of their consequences but, unlike utilitarians, he thought that the value placed on those consequences was an *objectively* identifiable amount of good.[2] Given the objective quality `good´, Moore´s fundamental questions were: (i) to make correct ethical judgements of what things possessed the quality `good´, and (ii) to identify the actions that caused `good´ things to come about. He concluded that we must follow certain rules which, according to him, represent the accumulated knowledge in our society that, in most cases, the rule results in the highest expected good. In other words, Moore´s conclusion was that the `best´ action was the one with the most likely good. In turn, this

implied that we should consider *both* the goodness of an action´s consequences and the relative frequency with which they occur. As Bateman (1988, p. 1100) notes, Moore `accepted that most of the rules adhered to in society did, in fact, *generally* cause the greatest good´. By contrast, Keynes and his friends were radically opposed to obeying rules in spite of the fact that they became members with Moore in the secret Cambridge society known as the *Apostles*. Many years later, in a piece titled `My Early Beliefs´ published in 1938, Keynes (1973*a*, p. 445) admitted that the large part played by considerations of probability in Moore´s ethics was a crucial contributory factor to his decision to study probability theory in the subsequent years.

In a series of notes Keynes wrote in 1905 titled *Miscellanea Ethica* he sought to attack Moore´s probabilistic argument for rules and chose to argue against Moore´s use of the *frequency* theory of probability (Bateman, 1988, p. 1101). In particular, in the piece titled `Ethics in Relation to Conduct´ Keynes sought to uncover the fallacies in Moore´s argument for rules. The alternative theory of probability he resorted to was one in which probability is an objective `degree of belief´ in a certain proposition rather than the relative frequency with which a certain phenomenon occurs. As Bateman (*op. cit*., 1102) recognizes, this theory of probability was a nascent version of the theory Keynes would propound some years later in *A Treatise on Probability*. According to him, the replacement of a frequency approach to probability by the notion of `degree of belief´ meant that:

`In Keynes´s theory of probability the good of each possible outcome would be multiplied by the degree of belief that one had in its occurrence, giving a product of good and degree of belief. The summation of these products would represent the expected good of an action and might be said to represent the degree of expected goodness of the action.

This alternative conception freed one from Moore´s "rules of Common Sense" since one was no longer obligated to act according to a frequency of [past] outcomes. Each case could now be considered on its own merits´ (Bateman, 1988, p. 1102).

Bateman (1988, p. 1103) also argues that Keynes followed Moore´s method of describing the `good´ in identifying the existence of a logical relationship as the basis for one´s `degree of rational belief´ in a proposition. Whereas Moore identified good as an objective concept that was indefinable in terms of anything else, Keynes (1920) also defined probability as an objective concept indefinable in terms of anything else.[3] By putting together his (objective) notion of probability as the `degree of rational belief´ and Moore´s (objective) notion of good as indefinable in terms of anything else, *Keynes concluded that it was possible to make expected utility calculations solely on the basis of our (objective) knowledge of both good and probability* and, unlike Moore, without having to resort to long-run relative frequencies of outcomes (Keynes, 1920, ch. 26). In other words, Keynes adopted Moore´s notion of good and built on it by developing a logical theory of probability. He then replaced the frequency theory of probability upon which Moore´s ethics was based by his own notion of probability as `degree of rational belief´. The end result was an ethics that differed markedly from Moore´s in the sense that following rules was not justified. Lastly, as Bateman (1988, p. 1104) notes, Moore came to accept in *Ethics* (Moore, 1912) Keynes´s definition of probability as the `degree of belief´ in a proposition.

2.2. Hume, Keynes & induction

Some scholars have suggested that Keynes wrote the *Treatise on Probability* to answer Hume (Fitzgibbon, 1988, p. 10; Davis, 1994, p. 142; Klant, 1985).[4] For instance, Klant (1985) argues that Keynes´s main intention in writing the *Treatise* was to solve the `problem of induction´ that had been set out by Hume (2006[1748]). According to Keynes (1920), Hume´s case against induction had not yet been answered: `Hitherto Hume has been master´ (Keynes, 1920, pp. 303-04). He thought that the logical system of Russell (and Whitehead) was constrained to exploring logical deductions and, hence, could not deal with the *non-deductive* method of reasoning that most people employ. As Andrews (1999, p. 2) remarks, `Keynes wished to construct an analytical structure for probable reasoning that would, at least to a certain extent, parallel that which Russell and Whitehead had produced for deductive reasoning´. This way, he would answer to Hume who concluded more than two hundred years before that: (i) it is not possible to ground probabilistic judgements in reason, and (ii) that the former are held only on the basis of repetition or `custom and habit´. Hume (2006[1748]) believed that *a priori* theorizing cannot tell us much about empirical relationships of cause and effect and that all knowledge we have of them comes from experience. Furthermore, he thought we do not experience causality relationships `directly´ but rather *infer* them from the observation that certain phenomena concur. Yet, he argued that in order to infer from a constant conjunction of phenomena that a causality relation actually exists requires to make the additional (and unwarranted) assumption that the `future must conform to the past´ (Hume, 2006[1748], p. 652). As we know, Hume argued that this assumption was *unprovable* in the sense that it was not liable to demonstrative proof. Nevertheless, he concluded that despite the impossibility of proving it, people constantly make decisions that are based on it due to the fact that they have no better alternative if they seek to pursue

their interests (Hume, 2006[1748], pp. 269-70). Hence, his profound skepticism about human rationality.

Keynes apparently found Hume´s argument against induction powerful although inconclusive: `Hume showed, not that inductive methods were false, but that their validity had never been established and that their possible lines of proof seemed equally unpromising´ (Keynes, 1920, p. 313). Unlike Hume, Keynes believed that there was a more robust foundation for probabilistic knowledge than `custom and habit´ and thought he had found it in the notion of probability as a logical relationship between two sets of propositions: premises and conclusions. In any case, his ulterior abandonment of the notion of probability that he had advocated in the *Treatise* led him to embrace a view of human rationality that was similar to Hume´s.

Next, Keynes´s *Treatise on Probability* addresses the question how knowledge is acquired. According to Keynes (*op. cit.*, pp. 3-4), knowledge is of two different forms: direct and indirect. First, direct knowledge is knowledge about (empirical) propositions concerning objects of direct acquaintance and where the propositions can derive from personal experience, sensorial perception or understanding. In any case, Keynes argues that the `true´ objects of knowledge and belief — in opposition to the objects of direct acquaintance — are not sensations, meanings, and perceptions but *propositions*. Keynes (*op. cit.*) denotes them as `primary propositions´. As he explains, the latter consist of propositions that capture our knowledge of the world and a further proposition that expresses a conjecture about the real world. For instance, given our current knowledge of the workings of the Eurozone and Greek history and politics, we think that it cannot be ruled out that Greece will eventually defect from the Eurozone. By contrast, indirect knowledge consists of `secondary propositions´ such as probability relations which are the result of arguments concerning `primary propositions´ and which may result either

from the application of `formal´ logic (i.e., demonstrative knowledge like mathematics) or `human´ logic (i.e., non-demonstrative knowledge).[5] In our example, the `secondary proposition´ consists of the probability that Greece will eventually leave the Eurozone. According to Keynes, theory construction consists of generating indirect knowledge in the form of `secondary propositions´. In any case, he argues that in all knowledge there is always some direct element so that `logic can never be made purely mechanical´ (Keynes, *op. cit.*, pp. 14-15) and it is on the basis of it that we seek to justify many further propositions or:

'That part of our knowledge which we obtain directly, supplies the premises of that part which we obtain by argument. From these premises we seek to justify some degree of rational belief about all sorts of conclusions' (*op. cit.*, p. 121).

Next, by the time Keynes wrote the *Treatise* he was mainly concerned to provide a logical foundation for induction, such as Whitehead and Russell (1925) had done for deduction. He was a student of probability theory before he became an economist and developed his ideas on probability prior to those on uncertainty. He was concerned with probability as the foundation for rational decision making rather than with descriptive statistics or with probability as a feature of the world. According to Carvalho (1998, p. 66), his main motive in writing the *Treatise* was to explain how a certain 'degree of belief' and the subsequent choice can be shown to be a *logical* consequence of a given set of premises and, hence, not merely a matter of the believer´s psychological features. Since he was reluctant to study the mechanisms through which the premises upon which decisions are based are generated, he skipped the issue *by assuming that the premises*

upon which the decision-making process is based were true knowledge (*op. cit.*). By so doing, Keynes was able to focus on deriving knowledge *by argument*. Consequently, the conclusion thus obtained was `rational´ to the extent that it did not depend on individual peculiarities but rather on criteria of consistency with `formal logic´.

Now, Keynes was aware that in many real-world circumstances neither certainty nor impossibility can be shown by `formal logic´. In all such cases, the relation between the premises and the conclusion was only *probable*. As noted in Carvalho (1988, p. 72), probability is approached in the *Treatise* as part of a learning process, that is, the larger is the (direct) evidence that accumulates in the form of premises, the more likely the conclusions obtained by argument are. This `cumulative´ view of learning is a logical implication of the inductive theory of knowledge Keynes had adopted. Indeed, it will become clear below that the probability theory of the *Treatise* is based on an *inductive* theory of knowledge. This is evident in chapter II and Part III (Induction and Analogy) of the *Treatise*. An example of this is the following quotation from the *Treatise*:

'The validity of the inductive method does *not* depend on the success of its predictions. Its repeated failure in the past may, of course, supply us with new evidence, the inclusion of which will modify the force of subsequent inductions. But the force of the old induction *relative to the old evidence* is untouched. The evidence with which our experience has supplied us in the past may have proved misleading, but this is entirely irrelevant to the question of what conclusion we ought reasonably to have drawn from the evidence then before us' (Keynes, 1920, p. 221).

Further, Keynes identifies knowledge with `certainty´ and, crucially, he assumes that knowledge (and, hence, certainty) is *obtainable*:

‘All knowledge, that is to say, which is obtained in a manner strictly direct by contemplation of the objects of acquaintance and without any admixture whatever of argument and the contemplation of the logical bearing of any other knowledge on this, corresponds to *certain* rational belief and not to a merely probable degree of rational belief’ (*op. cit*., p. 16).

Notwithstanding it, as David Hume showed long ago, direct knowledge cannot generate a set of `true´ premises that allow us to maintain a conclusion with certainty because our knowledge is always finite whereas the evidence is potentially infinite and, hence, never conclusive. Unfortunately, Keynes did not provide any further discussion about the epistemological status of the premises in the *Treatise* and the development of his ideas on decision-making tended to shift from probability to uncertainty.[6]

Next, Bateman (1991, p. 104) has argued that like Moore and Russell in the first decade of the 20th century Keynes `was committed to a very strict type of ontological realism; he believed that things and concepts have a real existence´. In this respect, he believed that there are propositions which describe the (non-physical) world and that the process of acquiring knowledge about it consisted of knowing these propositions. In turn, acquisition of knowledge was based on an inductive mechanism. A large part of the *Treatise* was thus devoted to the examination of the probabilistic nature of inductive arguments. Let us distinguish between the two types of inductive arguments Keynes identified. The first type is known as *universal induction* or arguments about invariable relations such as `all swans are white´ in which we seek to determine the correct `degree of belief´ in a given proposition that is either true or false. The other type of inductive argument is known as *statistical induction* and deals with arguments such as `ninety percent of swans are white´. Keynes devoted Part III (“Induction and analogy”) of the

Treatise to the analysis of universal induction and Part V ("Foundations of statistical inference") to the analysis of statistical induction.

Let us first address the case of universal induction. According to Keynes, every universal induction is made up of two elements: likeness and number. When we observe several instances of a phenomenon and make use of them to lend support for a universal induction, we rely on: (i) the fact that each instance exhibits the necessary likeness (e.g., all the observed birds are swans and they are all white), and (ii) the number of times we observe the *same* phenomenon. Keynes denoted these two elements of an inductive argument by *analogy* and *pure induction* respectively. He thought that pure induction was of very little use and that the correct way to make empirical arguments is through analogy. He then distinguished between positive and negative analogy (Keynes, 1920, p. 223). The former captures all the similar features shared by the observed instances whereas the latter refers to all their dissimilar features. According to Keynes (*op. cit*., p. 259), it is only an *increase* in negative analogy that enhances the validity of empirical arguments. Therefore, although the elements of an inductive argument are likeness (or analogy) and number, the latter is relevant only to the extent that it provides evidence about negative analogy.

As Bateman (1990, p. 366) notes, Keynes´s justification for inductive arguments depended crucially on two explicit assumptions: (i) the `law of uniformity of nature´, and (ii) the existence of `limited independent variety´ (Keynes, 1920, ch. 22). Firstly, the validity of inductive arguments depended on the ability to ignore *differences in time and space* among all observed instances. In other words, one needed to assume both that past observations would be valid in the future and that the generalizations made held regardless of our geographical location. Secondly, Keynes also believed that a positive *a priori* probability — which was a necessary condition for deriving the probability of a

generalization after *n* observations within his framework — could be assigned to a `primary´ proposition provided it could be assumed that there were a *finite* range of features which were relevant to the generalization. As Bateman (1990, p. 366) observes, although Keynes felt that these assumptions were plausible, he could not prove them and consequently could not offer a justification for induction.

Notwithstanding the previous conclusion, it has been suggested that Keynes´s discussion of statistical inductions served him to make the point that probabilities are not tantamount to relative frequencies. As Bateman (1990, p. 367) recognizes, Keynes believed that `the probability of observing a white swan was not necessarily the same as the percentage of swans which were known to be white´. More specifically, he believed that what one was trying to figure out was not what proportion of observed swans were white but what `degree of belief´ it was reasonable to hold in different statements about that proportion. He insisted that we should not confuse the past occurrence of a relative frequency with its future occurrence and that probabilities were `degrees of belief´ and not relative frequencies. In chapter 28 of the *Treatise* titled `The law of great numbers´ he traces the origin of the argument that after a large enough number of observations a stable pattern of frequencies will emerge and points to French mathematician Quetelet for the formulation of this law. Keynes concludes that `observation shows that some statistical frequencies are, within narrower or wider limits, stable. But stable frequencies are not very common, and cannot be assumed lightly´ (Keynes, 1920, p. 368). As we will see below, this rejection of a frequency interpretation of probability in the *Treatise* is a direct antecedent of his later claim, in the context of his criticism of Tinbergen´s pioneering work on the application of statistical methods to economics, that `economic data are not homogeneous through time´. Finally, an analysis of the foundations of the `law of great numbers´ is in chapters 29-31 of the *Treatise* where Keynes argues that it is

based on *unwarranted* assumptions about the distribution of the variables. The main source of Keynes´s disagreement with advocates of the `law of great numbers´ was that their method was *not* inductive:

`[Keynes] felt that instead of examining many instances and trying to *induce a pattern* from these, they had relied on undifferentiated observations and *assumptions about a pattern*. His complaint was not that they had used formulas or theorems but that (in his perception) these were used to circumvent the process of induction; instead of explaining how one could induce the likelihood of a frequency, these methods assumed the problem away´ (Bateman, 1990, p. 368).

Keynes´s alternative to the (relative) frequency theory of probability consisted of an *extension* of the elements of a universal induction to the case of statistical induction. In particular, he argued that in universal induction one observed a series of instances in varying circumstances and used this series to establish that a certain invariable relation (i.e., empirical regularity) held in many different circumstances. By contrast, in the case of statistical induction a series only represented one instance since one needed to count the relative frequency with which different characteristics appeared in the series. Thus, as Bateman (1990, p. 369) explains `in this case a "series of series" constituted the same thing as a series of observations in universal induction´. In particular, one would need to observe the relative frequency of occurrence of a certain feature (e.g., white swans) in many different series and increased the probability of the frequency´s future occurrence (e.g., that ten percent of all swans are white) by increasing the negative analogy (i.e., by observing that frequency in the widest array of circumstances) between series. In sum,

Keynes´s argument against early proponents of the frequency theory of probability was that their methods amounted to *pure* induction or *number* insofar as they did not take (negative) *analogy* into consideration (Keynes, 1920, pp. 445-46).

2.3. Probability

Following Gillies (2000), interpretations of probability can be divided into (i) epistemological, and (ii) objective. According to the former interpretation, probability is concerned with the knowledge or belief of human beings. In other words, probability measures `degree of belief´, `degree of rational belief´, etc. However, `epistemological´ interpretations of probability can themselves be either objective or subjective. Keynes, Ramsey, De Finetti, and Savage all adopted `epistemological´ approaches to probability theory. Notwithstanding it, while Keynes initially adopted (in his *Treatise*) an objective `epistemological´ interpretation, Ramsey, De Finetti and, Savage adopted a subjective `epistemological´ interpretation of probability.[7] By contrast, according to `objective´ interpretations of probability, the latter is a feature of the physical world which has nothing to do with knowledge or belief. In other words, human beings may know this probability or they may not but the probability is said to *exist* independently of whether it is known. Two well-known examples of `objective´ interpretations of probability are: (i) von Mises´s (1928) frequency theory of probability, and (ii) Popper´s propensity interpretation of probability[8] (Popper, 1959). The frequency theory of probability has been adopted by most mathematicians and consists in the description of the forms and properties of the randomness exhibited by the object of knowledge. An implication of it is that, if the stochastic processes are stable (or ergodic) then repeated observation of the object of knowledge allows us to acquire knowledge about its underlying structure. It is this feature of stochastic processes that Keynes apparently rejected and which made him

sceptical of the application of statistical methods in economics. Likewise, his scepticism about the possibility of economic data being `homogeneous through time´ explains, to a large extent, his view of theoretical models as instruments for `explaining´ economic phenomena rather than for generating `forecasts´.

The *Treatise on Probability* (Keynes, 1920) is, along with the work of Jeffreys (1939) and Carnap (1950), part of the `logical´ tradition in the theory of probability. In this tradition, probability is conceived as a *logical* relation between a conclusion and certain evidence. In Keynes´s view, logic had always considered such relationships but it had only dealt with the cases in which one´s information implied that a proposition was true or false. His project in the *Treatise* was to extend logic to those cases — the most common in science and life — where we are uncertain about a proposition given the evidence available. This feature is stressed in Runde (1990, p. 277) who explains that 'unlike the frequency theory of probability in which probability is interpreted as a property of the physical world, Keynes treats probability as a property of the way individuals think about the world'. In the same vein, Carvalho (1988, p. 71) points out that in the frequency theory of probability 'the accumulation of knowledge does not change probabilities, because randomness is a feature of the object of knowledge, not of knowledge itself´. By contrast, Keynes´s notion of probability implies that a probability proposition may *change* when the evidence upon which reasoning is based changes.

Next, in the *Treatise* Keynes draws a sharp distinction between the certainty of the conclusions of demonstrative logic and the inconclusiveness of non-demonstrative logic on which scientific (and non-scientific) arguments are typically based. As we have explained above, he defends the *real* existence of a logical relation between two or more propositions in those cases where it is impossible to construct a demonstrative argument from one to the other. He denotes this logical relation as the *probability relation* or

argument. Therefore, a conclusion *a* is related to a set of premises *h* by virtue of a probability relation. This is written as *a/h=p*, where *p* is the 'degree of rational belief' that the probability relation between *a* and *h* justifies. In turn, Keynes (1920) denotes this probability relation as a 'secondary proposition' which, as we explained above, constitutes a logical statement about the 'primary proposition' *a*. The probability of a conclusion *a* is always relative to a set of evidence *h*. If the relation between *a* and *h* were tautologous or true then *p=1* whereas if it were contradictory or false then *p=0*. If the evidence did not guarantee the truth or falsehood of the proposition, there existed a `degree of rational belief´ between *0* and *1* in the propositions´ truth. The discovery of new evidence *h1* will not affect the validity of the probability relation between *a* and *h*, but will yield a new one *a/hh1*. This way, Keynes extended Russell´s logic of certainty (true and false propositions) to uncertainty (i.e., to more or less probable propositions).

Now, Keynes´s approach to probability might be interpreted as subjective since the amount of information and reasoning power may vary across individuals. However, Keynes stresses the objective component which he denotes as the 'degree of belief' it is *rational* to hold in a certain hypothesis given the evidence available (Keynes, 1920, p. 4). As Cottrell (1993, p. 27) notes, there is a *Platonic* element in Keynes´s interpretation of probability in the sense that probability relations exist in a certain logical space and are assumed to be perceived by individuals via intuition. This implies that, if opinions that are based *on the same evidence* differ, it follows that somebody´s perception must be wrong. In other words, Keynes´s approach to probability implies that probabilities are *unique* and have the same value for all the individuals who happen to possess the *same* information. By contrast, in subjective theories of probability the probabilities will take whatever value is assigned by the individual using them. This means, for instance, that in the subjective theory of probability originally propounded by Ramsey (1926) and De

Finetti (1937), any two individuals with the same information can assign *different* numerical values to the probability of a certain proposition whereas, in Keynes´s theory of probability, two individuals with the same information will assign the *same* value to the probability of a proposition. Crucially, the objectivity of probabilities in Keynes´s theory is not based on inter-subjective agreement among individuals — as in the inter-subjective approach propounded in Gillies (2000) — but on the real existence of logical Platonic-like probability relations. As Lawson (1988, p. 44) concludes, the difference between Keynes´ approach to probability and the approach adopted by Ramsey (1926), De Finetti (1937) and, later on, Savage (1951) is that 'although both approaches accept that probability exists only on the level of knowledge or opinion Keynes believes that probability is objectively determined'.

2.4. Rationality

Both Bateman (1987, pp. 101-102) and Andrews (1999, p. 6) make the point that Keynes´s argument for the objectivity of probabilities is related to the inductive theory of knowledge expounded in Parts III and V of the *Treatise*. According to Keynes, there is an essential connection between the inductive methods and the notion of probability he presents in the *Treatise*. In particular, he argues that an induction is just a *particular* application of his theory:

`An inductive relation affirms, not that a certain matter of fact is so, but that relative to certain evidence, there is a probability in its favour´ (Keynes, 1920, p. 254).

That is, Keynes believed that probability is a `logical´ relationship between two propositions: (direct) evidence and conclusion. Further, *he also believed that decisions are based on probabilistic judgements. To the extent that human behaviour is grounded upon objective probability relations, the former is `rational´*. Thus, unlike Hume, who believed that human behaviour was grounded on `custom and habit´, he concluded that rational probability is the `guide of life´ (*op. cit.*, p. 369). As Bateman (1991, p. 105) writes, `the end result of all this was a conception of probability which relied on intuition as the basis of rational choice´. Specifically, individuals were *assumed* to know the `correct´ probabilities of the various outcomes — given the evidence available — and to weight the value of each possible outcome by its probability. In turn, Keynes´s argument that probabilities are `rational´ is based on the idea that they had been *inferred* correctly from the available information through inductive methods (Bateman, 1987). Therefore, the whole process was *rational*. Be that as it may, we will see below that his view of rationality underwent a dramatic change in the aftermath of his acceptance of Ramsey´s criticism of logical probability relations (Ramsey, 1926) and those changes pushed him back towards the views on human nature held by Hume.

In terms of the conceptual framework we used in the previous chapter, we may assert that the logical probability relations Keynes devised in the *Treatise* satisfy means-rationality and beliefs-rationality (Hamlin, 1986).[9] Yet, this is not to deny the possibility that realized outcomes may differ from *ex-ante* expectations. To be sure, an individual who makes decisions according to Platonic-like probability relations may well obtain a result which differs markedly from the expectation she held with a certain `degree of rational belief´ yet such unexpected outcome cannot be said to be a mistake but rather it is a mere consequence of the fact that, in Keynes´s *Treatise*, knowledge is probabilistic and, therefore, liable to error. Further, the theory Keynes developed in the *Treatise* can

be interpreted as being both a positive and a normative theory, that is to say, as being a description of human behaviour as well as a `guide´ to it. To be sure, Keynes´s assertion in the *Treatise* that human decisions are based upon probability judgements and that the latter are `rational´ insofar as they are built upon `correct´ inductive inferences makes it unequivocal that he conceived of probabilistic judgements as descriptions of individual behaviour. By the same token, if human behaviour is `rational´ it then follows that it is `desirable´ that the former be based upon logical probability relations.

Next, O´Donnell (1990, p. 257) argues that part of the deeper analytical structure of the *Treatise* is constituted by two distinct domains of analysis. The first domain is the *determinate* domain, the ruler of which is *strong rationality*. As in Simon´s well-known notion of `substantive rationality´ (Simon, 1976), within the domain of strong rationality determinate answers are available to the questions posed and it is assumed that reason is capable of providing the answers. Our previous discussion about Keynes´s notion of probability as `degree of rational belief´ in the *Treatise* and its accompanying notions of `primary´ and `secondary´ propositions do correspond to O´Donnell´s notion of *strong rationality*. The other domain identified in O´Donnell (*op. cit.*) is the *indeterminate* one and is associated to the notion of *weak rationality*. As O´Donnell (*op. cit.*) argues, in the *indeterminate* domain 'probabilities are unknown for want of reasoning power; it is theoretically impossible to compare probabilities; weights are also non-comparable; and agents are thus unable to combine probabilities and values where either or both are non-numerical'. According to O´Donnell (1990, p. 258), the theory of the *Treatise* embraces both domains and, although the development of the determinate domain was Keynes´s major concern in and prior to the *Treatise*, the existence of an indeterminate domain was recognised albeit relegated to a subordinate position. Nevertheless, he adds that it is certainly the indeterminate domain that Keynes sought to develop in his later economic

writings.[10] In turn, this serves him to insist on the `continuity´ thesis, that is, the claim that there is a basic continuity in Keynes´s conceptual framework between the *Treatise on Probability* and the *General Theory*) against advocates of the `non-continuity´ thesis, i.e., that there is a substantial change or a `break-up´ in Keynes´s thought between the *Treatise* and his later economic writings (Andrews, 1999; Bateman, 1987, 1990, 1991; Davis, 1994).[11]

Be that as it may, O´Donnell (1990) insists that the absence of strong rationality in Keynes´s conceptual framework in his later economic writings is not tantamount to 'irrationality'. Rather, he contends that individuals will adapt to the new circumstances and develop strategies to 'get by'. He groups such strategies under the heading of *weak rationality*; in the indeterminate domain rational individuals know that weakly rational procedures result from the lack of robust foundations in certain or probable knowledge and, hence, they have low and easily shaken, degrees of confidence in such procedures. Further, he argues that Keynes observes that despite being uncertain of the future people know certain facts of the existing situation as well as ways of 'getting by' in the absence of knowledge of the relevant (numerical) probabilities. Such ways of 'getting by' bear a resemblance to Simon´s notion of 'procedural' rationality (Simon, 1976). For instance, in the context of investment decisions, Keynes argues that investors often 'endeavour to fall back on the judgement of the rest of the world which is perhaps better informed' (Keynes, 1937, p. 214). However, this interpretation of Keynes´s notion of rationality as implying two domains has been criticized in Bateman (1991). He refers to O´Donnell´s (1990) notion of *weak rationality* as follows:

`Here all of Keynes´s careful argument about probability as the "guide of life" comes tumbling down. With one deft stroke, the people who follow social conventions

and conventional rules of behaviour are out on a par with those who have the full ability to recognise (and act upon) Platonic probabilities´ (Bateman, 1991, pp. 108-9).

According to Bateman (*op. cit.*), O´Donnell´s interpretation of Keynes´s *Treatise* has two problems: (i) that Keynes never mentioned a second theory of rationality in his *Treatise*, and (ii) that his many references to `irrational´ behaviour in his later economic writings suggests that, far from having a *dualistic* theory of human rationality, he made `irrationality´ the basis of his macroeconomics. Although we do not subscribe the latter appraisal of the theory of human behaviour underlying Keynes´s *General Theory*, we share Bateman´s argument that presenting Keynes´s framework in the *Treatise* as made up of two different domains of analysis is inappropriate in that it places more emphasis on the existence of continuity in Keynes´s thought than it seems to be justified by the textual evidence. Having said that, we believe it is in the following ideas that appear in the *Treatise* where there is a clear connection between his previous work on probability theory and his later economic writings: (i) the idea that some probabilities may not be measurable, (ii) the `weight of an argument´, and (ii) the illegitimacy of the frequency approach to probability in economics.[12] We address the first two issues in what follows.

2.4.1. Uncertainty

Keynes does not explicitly define uncertainty in the *Treatise*. Uncertainty seems to be related to the absence of certainty but it is a particular lack of certainty. To be sure, Lawson (1987, p. 953) argues that in the *Treatise* Keynes identifies three different types of probability relations: (i) where a probability relation is numerically indeterminate and, possibly, not comparable to other probability relations, (ii) where probabilities are numerically determinate and take on values between unity and zero, and (iii) where

probabilities take the value of either unity or zero. According to Lawson (1985, p. 911), Keynes associates the first type to a situation of 'uncertainty' and the third type to a situation of 'certainty'. Next, focusing on the first type of probability relation, Keynes (1920) identifies two scenarios in which `direct´ knowledge and, hence, certainty of a secondary proposition or probability relation concerning a given primary proposition *a* based on a certain type of evidence *h* is absent. The first scenario is when the secondary proposition *a/h=p* is *unknown* because the weakness of our reasoning and computing power prevents us from knowing what the probability relation is in spite of the (direct) evidence justifying a certain degree of knowledge (*op. cit.*, p. 34). Let us refer to the type of uncertainty that stems from this first scenario as *epistemological*. The second scenario Keynes identifies is when probability relations *can be said not to exist or when they are numerically immeasurable or indeterminate* because there is no basis possible, given the existing evidence, for determining the numerical probability of the secondary proposition. Let us denote the resulting type of uncertainty as *ontological*.

Now, Lawson (1985, p. 914) insists that it is the second scenario where `direct´ knowledge of a secondary proposition is absent because there is no basis upon which to generate any (numerical) probability that corresponds to a situation of 'uncertainty' in Keynes. In other words, according to Lawson, 'uncertainty' is not a situation in which the probability relation is known and where the primary proposition *a* relative to the evidence *h* yields a numerical probability that is *less* than unity. Neither is it a situation in which an existing probability relation cannot be grasped due to our limited cognitive powers. Rather, he insists that 'uncertainty' in the *Treatise* arises when the probability relation that concerns a primary proposition *a* is *numerically indeterminate*. In short, Keynes´s notion of 'uncertainty' is *ontological* albeit, in some circumstances, the source of uncertainty may well be epistemological. According to Lawson, this interpretation of

Keynes´s notion of uncertainty in the *Treatise* is consistent with Keynes´s use of the same term in the *General Theory* and in his 1937 *Quarterly Journal of Economics* paper where he says that 'very uncertain' is not the same as 'very improbable' (Keynes, 1937, pp. 213-14).

2.4.2. The 'weight of an argument'

Let us now address Keynes´ notion of the 'weight of an argument'. Runde (*op. cit*, p. 133) denotes Keynes´s theory of expectations formation as a 'two-tier theory of belief' whereby probability is at a first level whereas weight is at the second level. The former represents a measure of our `degree of belief´ in a hypothesis *relative* to some set of evidence whereas the latter constitutes a measure of the `degree of completeness´ of the evidence on which that belief is based.[13] According to Runde (1990), Keynes alludes to the notion of weight in the *Treatise on Probability* in, three different ways.[14] According to the first conception of weight (weight1), 'one argument has more weight than another if it is based on a greater amount of relevant evidence' (Keynes, 1920, p. 77). The second conception of weight that Runde identifies (weight2) appears in Chapter 6 of the *Treatise* where weight is defined as the 'balance of the absolute amounts of relevant knowledge and relevant ignorance on which a probability is based' (*op. cit.*, p. 280). To be sure, if we denote relevant knowledge and relevant ignorance by *Kr* and *Ir* respectively and weight by *V*, then the latter may be expressed as:

$$V(a/h) = Kr/Ir \qquad (1)$$

The third notion of `weight´ appears in Chapter 26 of the *Treatise* where Keynes makes it very clear that 'the probability of a hypothesis is only one of the things to be determined and taken account of before acting on it' (Keynes, 1920, p. 307). It is clear

then that, as long as probabilities are used as a guide to action, the other element when making decisions is weight. The third notion of weight (weight3) identified by Runde is the 'degree of completeness of information on which a probability is based' (*op. cit*., p. 313). As Runde (*op. cit*.) remarks, if knowledge, information, and evidence are taken as being synonymous terms, as Keynes apparently does, then the 'degree of completeness' of the information on which a probability is based might be expressed as:

$$V(a/h) = Kr / (Kr + Ir) \qquad (2)$$

From expressions (1) and (2) we have that, as the amount of (absolute) relevant knowledge increases, the probability of proposition *a* may either rise or fall, depending on whether the new evidence is favourable or unfavourable.[15] Thus, provided that *Ir* does not increase by more than *Kr*, the weight of the new probability will be *higher* than it was before. Yet, and following this interpretation, *the discovery of new evidence may lead to a decrease in weight.* This will be the case if, for instance, *Ir* increases by more than *Kr*. Runde (*op. cit.*, p. 283) admits that this possibility is not explicitly contemplated in the *Treatise* and it may contradict some of Keynes´s statements. Notwithstanding it, he provides a number of examples taken from the *Treatise* itself which seem to imply the consequence that, if new evidence is discovered so that our (subjective) perception of ignorance increases (i.e., the degree of completeness of the information decreases) either because more options have now become available or because a larger array of possibilities can now be considered, then weight may *decrease* even if our (absolute) knowledge increases. In other words, according to Keynes, weight measures the 'well groundedness' of the argument and, hence, the confidence we place on it.[16] Low weight means weaker foundations for decision making and low degrees of confidence whereas high weight implies strong foundations and higher degrees of confidence, even though the probabilities involved in both scenarios may be perfectly correct.

Now, whereas the first notion of weight we have identified above is objective, expressions (1) and (2) above corresponding respectively to the second and third notion of weight identified in Runde (*op. cit*.) are clearly subjective. In the case of the second version presented above, the 'balance of the absolute amounts of relevant knowledge and relevant ignorance on which a probability is based' (*op. cit*., p. 280) can only be a subjective perception because the absolute amount of factual evidence that is potentially available is *infinite* and, hence, no observer can `objectively´ know what the balance of relevant knowledge and ignorance is.[17] Likewise, in the third notion of weight identified in Runde the `degree of completeness of information on which a probability is based' can only be a subjective perception for the same reason that the amount of information is infinite and, hence, there is no way to know what the *actual* `degree of completeness´ of the information we possess is. In other words, there is no way to know *how* ignorant we are about something we ignore. This explains why the notion of weight appears to be confusing for Keynes himself as Cottrell (1993, p. 36) notes.[18] Further, this suggests that the 'two-tier theory of belief' that Runde (*op. cit*.) refers to does not only comprise probability or `degree of rational belief´ and weight but really comprises an apparently *objective* notion of probability coupled to a *subjective* notion of `weight´. We believe it is for this reason that Keynes easily connected the notion of `weight´ of the *Treatise* to the twin notions of `confidence´ and `liquidity preference´ in the *General Theory*.[19] In particular, the link is facilitated by the fact that two of the three versions of the notion of `weight´ displayed in the *Treatise* are subjective perceptions. The subjective nature of `weight´ also goes some way, according to us, towards explaining Keynes´s capitulation to Ramsey´s critique (Ramsey, 1978[1931]) when the latter argues that Keynes´s logical probability relations do not exist and that probability is *subjective*.

2.5. Some final remarks on the *Treatise on Probability*

We have shown that Keynes´s *Treatise on Probability* constitutes, at least partly, an answer to Hume by providing both an objective epistemological theory of probability and theory of rational decision-making based on an inductive theory of knowledge. As to the former, Keynes suggests that, although we cannot acquire `certain´ knowledge we can nevertheless attain probabilistic knowledge. However, if we can acquire `objective´ probabilistic knowledge by relying on inductive methods, the basis for both a `rational´ theory of decision-making under uncertainty and an answer to Hume have been laid out. Notably, Keynes *assumes* that individuals know *via* intuition the `correct´ probabilities of all possible outcomes — given the evidence accumulated — and weight the value of each possible outcome by its `correct´ probability. We have suggested above that this `embryonic´ theory of decision-making under uncertainty faces (at least) two problems: (i) the objectivity or `Platonic´ nature of logical probability relations and, arguably, (ii) our claim that the notion of the `weight of argument´ is *subjective*. The `objectivity´ of *a priori* probability relations was the focus of Ramsey´s critique of Keynes´s *Treatise* and will be addressed in the following section. As to our claim that the `weight of argument´ is subjective, we believe that it highlights the insurmountable conceptual difficulties of constructing a theory of human behaviour under uncertainty grounded, as Schumpeter (1984) would put it, upon the `rationality of the observer´.

According to us, the main conceptual difficulty the notion of *a priori* probability relations encounters is that the absolute amount of empirical evidence economic agents may come across is potentially *infinite*. In the frequency approach to probability this is dealt with by assuming that relative frequencies converge to a certain (constant) value as the number of observations increases. However, Keynes rejected the validity of the frequency approach to probability in the social sciences. As a result of it, he had to rely

on Platonic-like or *a priori* probability relations to justify their `objectivity´. Otherwise, how could he argue that such *a priori* probability relations are `objective´ if the amount of evidence that is potentially available and the different types if circumstances in which such evidence can be found are potentially infinite? As we will see below, the edifice laboriously constructed by Keynes in the *Treatise* crumbled irremediably when Ramsey criticised it. Keynes capitulated in the aftermath of it (Keynes, 1973*a*, pp. 338-39). The upshot of it is the apparent adoption by Keynes of a subjective epistemological theory of probability which would eventually pave the way to the final adoption of a theory of human nature similar to Hume´s and an (implicit) recognition of his inability to provide a solution to Hume´s `problem of induction´.

3. Ramsey´s critique of the *Treatise* and his theoretical framework

Like Keynes, Ramsey grew up in Cambridge, was a student of both Moore and Russell, an Apostle and a contributor to the *Economic Journal*. In spite of his early and unexpected death in 1930 when he was only 26, he made long-lasting contributions to probability theory, philosophy, and economics. For instance, it is widely recognized that Ramsey´s critique of the *Treatise on Probability* in his `Truth and Probability´ (Ramsey, (1978[1931]) constitutes, together with De Finetti´s (1937), the foundation for Savage´s subjective expected utility theory (Savage, 1954), which is still the standard approach to analysing decision-making in economics.[20] Before we present Ramsey´s critique of the *Treatise* let us note that in the foreword of the above-mentioned essay Ramsey suggests that there may be two legitimate approaches to probability:

`Probability is of fundamental importance not only in logic but also in statistical and physical sciences, and we cannot be sure beforehand that the most useful interpretation of it in logic will be appropriate in physics also. Indeed, the general

difference of opinion between statisticians who for the most part adopt the frequency theory of probability and logicians who mostly reject it renders it likely that the two schools are really discussing different things, and that the word "probability" is used by logicians in one sense and by statisticians in another´ (Ramsey, (1978 [1931], p. 59).

Thus, as with Keynes´s (implicit) acceptance some years after the publication of the *Treatise* of the legitimacy of the frequency theory of probability in the context of the natural sciences, Ramsey suggests that, at least in physics, the requirements for the use of the frequency theory of probability are fulfilled. Therefore, in both cases (Keynes and Ramsey) there is a clear *duality* in the theory of probability. We will return to this issue below when we address Keynes´s ideas on the methodology of economics.

3.1. Ramsey´s critique

We now address the essence of Ramsey´s critique of the *Treatise on Probability*. As we will see below, this critique had a dramatic impact on the views Keynes adopted on probability theory and the methodology of economics at least from 1932 onwards and, hence, it is an indispensable ingredient in any attempt to understand Keynes´s later economic writings. In a much-quoted paragraph, Ramsey writes:

`But let us now turn to a more fundamental criticism of Mr. Keynes´ views, which is the obvious one that there really do not seem to be any such things as the probability relations he describes. He supposes that, at any rate in certain cases, they can be perceived; but speaking for myself I feel confident that this is not true. I do not perceive them, and if I am to be persuaded that they exist it must be by argument; moreover I shrewdly suspect that others do not perceive them either, because they are able to come

to so little agreement as to which of them relates any two given propositions´ (Ramsey, (1978[1931], p. 63).

In other words, Ramsey denies that there are `objective´ logical relationships to which our `degrees of belief´ refer and insists that we simply have *subjective* `degrees of belief´ in different propositions. As Bateman (1987, p. 106) explains, Ramsey´s critique implies that, if two individuals possess exactly the same information but they exhibit different `degrees of belief´ in a certain proposition, this does not imply — unlike in Keynes´s *a priori* probability relations — that one of them is necessarily `irrational´ or wrong. Rather, the dissimilarity in their evaluations implies that these two individuals have made assessments of the likelihood of different possible outcomes that reflect their own *subjective* `degrees of belief´. Ramsey coined the term `human logic´ to denote the subjective nature of the assessments people make and to emphasize the absence of an *a priori* probability relation. The obituary containing Keynes´s answer to Ramsey was published in the *Economic Journal* in 1931, a year after Ramsey died, and leaves little doubt about his capitulation:

`Ramsey argues, as against the view which I had put forward, that probability is concerned not with objective relations between propositions but (in some sense) with degrees of belief, and he succeeds in showing that the calculus of probabilities simply amounts to a set of rules for ensuring that the system of degrees of belief which we hold shall be a *consistent* system. Thus the calculus of probabilities belongs to formal logic. But the basis of our degrees of belief — or the *a priori* probabilities, as they used to be called — is part of our human outfit, perhaps given us merely by natural selection,

analogous to our perceptions and our memories rather than to formal logic. So far I yield to Ramsey — I think he is right´ (Keynes, 1973*a*, pp. 338-39).

Therefore, while Keynes had advocated an objective epistemological theory of probability in the *Treatise*, he was now willing to accept a subjective epistemological theory of probability. Furthermore, Keynes was also willing to accept that the role that Ramsey assigned to the calculus of probabilities was to ensure the *consistency* of these subjective `degrees of belief´ (Bateman, 1987). What has gone largely unnoticed in the academic literature, however, is that when referring to the source of subjective `degrees of belief´, Keynes hints that they may have been `given us merely by natural selection, analogous to our perceptions and our memories rather than to formal logic´.[21] Although Keynes, to the best of our knowledge, does not develop this idea elsewhere, he ponders the possibility that our expectations are determined by a `natural´ selection mechanism, i.e., through a trial and error-elimination process similar to Popper´s evolutionary theory of knowledge and learning (PTKL). Be that as it may, in the same quotation he seems to stick to an inductive theory of knowledge:

`But in attempting to distinguish "rational" degrees of belief from belief in general [Ramsey] was not yet, I think, quite successful. It is not getting to the bottom of the principle of induction merely to say that it is a useful mental habit´ (Keynes, 1973*a*, p. 339).

As Bateman (1987, p. 107) suggests, while accepting the subjective nature of our `degrees of belief´, Keynes implies that consistent subjective `degrees of belief´ are *not* rational. Rationality in the *Treatise* stems from the fact that probabilities are assumed to

have been *inferred* correctly from the evidence available. By contrast, the rationality of probabilities in Ramsey´s framework stems from their *consistency* according to the rules of calculus of probabilities.[22] In a similar vein, Andrews (1999, p. 8) argues that Keynes believed that Ramsey´s subjective theory of probability failed to distinguish `rational´ belief from belief in general. The reason is that Keynes continued to adopt an inductive theory of knowledge and, hence, to associate `rationality´ with probabilistic knowledge. Notwithstanding it, Keynes eventually adopted Ramsey´s (as well as Hume´s) view of rationality. Ramsey (1978[1931], p. 96) argues that `the human mind works essentially according to general rules and habits´, denotes `human logic´ as the `logic of truth´ or `inductive logic´, and states that `induction is such a useful habit, and so to adopt it is reasonable´ (*op. cit*., p. 100). However, this is remarkably similar to Hume´s approach to rationality. Indeed, Ramsey explicitly makes the connection between rationality and induction when he refers to Hume:

`Hume showed that it [induction] could not be reduced to deductive inference or justified by formal logic. So far as it goes his demonstration seems to me final; and the suggestions of Mr Keynes that it can be got round by regarding induction as a form of probable inference cannot in my view be maintained. But to suppose that the situation which results is a scandal to philosophy is, I think, a mistake…We are all convinced by inductive arguments and our conviction is reasonable because the world is so constituted that inductive arguments lead on the whole to true opinions´ (*op. cit.*, pp. 98-99).

Thus, Ramsey equates `reasonable´ behaviour with the use of inductive methods and, in this way, he seems to subscribe Hume´s conclusion that, despite the absence of an `inductive logic´, there is a nevertheless a `human logic´ that is essentially inductive, i.e.,

individuals make use of induction to make decisions. We interpret his insistence on induction as a kind of pragmatism:

`We all agree that a man who did not make inductions would be unreasonable: the question is only what this means. In my view it does not mean that the man would in any way sin against formal logic or formal probability; but that he had not got a very useful habit, without which he would be very much worse off, in the sense of being much less likely to have true opinions´ (*op. cit.*).

What is important for our purposes is that, as we noted above, Keynes not only accepted Ramsey´s critique of his notion of objective logical probability relations but he ended up adopting by 1938 a theory of human behaviour similar to both Ramsey´s and Hume´s.[23] In *My Early Beliefs* (Keynes, 1973*a*, pp. 433-51), Keynes recalls the views of human nature held by the `Old Bloomsbury´ and charges himself and his friends at the 1938 Memoir Club meeting with holding ideas that were `disastrously mistaken´ in the sense that it presupposed the idea that:

`The human race already consists of reliable, decent people, influenced by truth and objective standards, who can be safely released from the outward restraints of convention and traditional standards and inflexible rules of conduct, and left, from now onwards, to their own sensible devices, pure motive and reliable intuitions of the good´ (*op. cit*., p. 447).

The alternative view he adopts is as follows:

`We were not aware that civilisation was a thin and precarious crust erected by the personality and the will of a very few, and only maintained by rule and conventions skilfully put across and guilefully preserved´ (*op. cit.*, p. 447).

The adoption by Keynes in his later economic writings of a *conventional* theory of beliefs and human behaviour is, for most purposes, tantamount to Hume´s view that individual behaviour is grounded on `custom and habit´. In other words, conventions or `custom and habit´ are two examples of factors which affect human behaviour *whose institutionalization can be ascribed ultimately to the fact that repeated observation (i.e., induction) has highlighted that their adoption yields desirable results*. The paradox is that Keynes´s main intention in the *Treatise* was to answer both Hume (1978 [1739-40]) and Moore (1912) by propounding a solution to the `problem of induction´ and an ethics that were not based on the adoption of social rules. Notwithstanding it, his capitulation to Ramsey implied, as we will see below, that he ended up adopting a theory of human behaviour based on the reliance by individuals on conventions where any pretension to Platonic-like `objective´ rationality was abandoned and replaced by a modest claim to `reasonable´ behaviour.

3.2. Ramsey´s theoretical framework

The rest of this section is devoted to exploring other aspects of Ramsey´s essay *Truth & Probability* which may also have a bearing on the methodology of economics beyond the above-mentioned impact on the development of probability theory. More precisely, we will argue below that Ramsey (1978[1931]) introduces a methodological principle that exhibits a remarkable similarity to Popper´s `Rationality Principle´. Since Keynes (1973*a*, pp. 433-51) read Ramsey´s essay, it is quite likely that the approach to

modelling behaviour in the social sciences that Ramsey sketches in that piece had an impact on Keynes´s thought. We reproduce below the crucial textual evidence in *Truth & Probability*:

`In order therefore to construct a theory of quantities of belief which shall be both general and more exact, I propose to take as a basis a general psychological theory, which is now universally discarded, but nevertheless comes, I think, fairly close to the truth in the sort of cases with which we are most concerned. I mean the theory that *we act in the way we think most likely to realize the objects of our desires*, so that a person´s actions are *completely* determined by his desires and opinions. This theory cannot be made adequate to all the facts, but it seems to me a useful approximation to the truth… I only claim for what follows approximate truth, or truth in relation to this artificial system of psychology which like Newtonian mechanics can, I think, still be profitably used even though it is known to be false´ (Ramsey, (1978[1931], p. 75, emphasis added).

There are several ideas in the previous paragraph which deserve our attention. First, the *general psychological theory* Ramsey propounds, `the theory that we act in the way we think most likely to realize the objects of our desires´ is remarkably similar to Popper´s *RP* according to which `our actions are adequate to our problem-situations as we see them´ (Popper, 1994, p. 181). Notably, Popper says explicitly that he views his *RP* as the animating part of any social theory just as the laws of motion of planets are an integral part of Newton´s gravitational theory (Popper, 1994, p. 177) whereas Ramsey makes an analogy between his general psychological theory and Newtonian mechanics and adds that both theories are `approximately´ true. In this respect, both Ramsey and Popper recognize that the principle they propound is either a good enough (Popper) or a

useful (Ramsey) approximation to the `truth´ and, for both of them, the status of the principle is *methodological*. That said, there are, at least, two differences between these two versions of the principle. First, Ramsey emphasizes the uncertainty surrounding the consequences of our actions when he proposes that we behave in the way we think *most likely* to fulfil our desires, whereas the presence of uncertainty in Popper´s *RP* is more subtle in that it is fully absent from the `objectivist´ version (*RPo*) but is present in the `subjectivist´ version (*RPs*) insofar as there may be a `wedge´ between the objective P-S and agents´ view of it. Second, Ramsey (*op. cit*.) stresses the purposefulness of human action and then argues that the latter is *completely* determined by our desires and beliefs without any consideration being paid to the P-S, whereas Popper makes P-S the centre of his proposal by stating that `our behaviour is adequate or appropriate to the P-S we face´ and, hence, he implicitly denies that human behaviour is completely determined by desires and beliefs.

These ideas are depicted in Table 2 below which shows a tentative classification of probability theories and the resulting approaches to rationality. Keynes´s approach to probability and rationality in the *Treatise* is shown in the first row. As it turns out, the theory that emerges can be interpreted as being both a positive and normative approach to rationality. As we suggested above, Keynes´s approach to rationality in the *Treatise* implies fulfilment of means-rationality and beliefs-rationality. Next, both Ramsey and Keynes (in the wake of *Truth and Probability*) adopt a *dualistic* approach to probability in the sense that they believe that the adoption of the frequency approach to probability is legitimate in the natural sciences but inappropriate in philosophy (Ramsey) or in the moral sciences (Keynes).[24] Furthermore, as we showed above, the main implication of Ramsey´s critique of the *Treatise* is Keynes´s acceptance of the notion that probabilities are subjective. In turn, we believe the former implies that any normative pretension was

abandoned by Keynes thereafter since: (i) the adoption of a subjective epistemological probability theory cannot imply fulfilment of `beliefs-rationality´, and (ii) fulfilment of means-rationality can only be understood as implying that behaviour is *consistent* with one´s desires and beliefs. Thus, we may characterize the theoretical framework Keynes adopted in the aftermath of *Truth & Probability* as being essentially `positive´.

Probability Theory		**Nature of probability theory**	**Status of theory**	**Type of rationality**	**Popper´s rationality principle**
Keynes´s *Treatise on Probability* (1952[1920])		Objective & epistemological	Positive & normative	Means-rationality & beliefs-rationality	—
Ramsey´s critique of Keynes´s *Treatise* (1931)	Philosophy	Subjective & epistemological	Positive	Ends-rationality, means-rationality & beliefs-rationality	Subjectivist version
	Frequency theory (physics)	Objective & aleatory	Positive	—	—
Keynes´s later economic writings (1936, 1937)	Frequency theory (natural sciences)	Objective & aleatory	Positive	—	—
	Conventional theory (moral sciences)	Inter-subjective & epistemological	Positive	Ends-rationality & means-rationality	Subjectivist version

Table 2. Classification of probability theories in Keynes´s and Ramsey´s work

Similarly, Ramsey´s framework can be interpreted as being a positive approach. It is essentially a `logic´ for thinking about human behaviour under the premise that the latter is purposeful and consistent with one´s objectives and beliefs. As such, Ramsey´s framework implies fulfilment of both `ends-rationality´ and `means-rationality´ but not `beliefs-rationality´.[25] Unlike in Savage´s (1954) framework, Ramsey does not assume

that the individual knows the consequences of adopting each possible course of action. Therefore, and unlike Savage´s framework, the absence of `beliefs-rationality´ prevents Ramsey´s framework from representing a normative theory. Further, the `logic´ Ramsey (1978[1931]) propounds to explain human behaviour and which was reproduced in the previous footnote implies that his approach represents an instance of what we defined in the previous chapter as *RPs* in the sense that the theoretician reconstructs P-S as `she believes the agent believes it is´. Though we do not wish to dwell on it, the adoption by Savage (1954) of Ramsey´s subjective epistemological interpretation of probability was coupled to the development of a framework where agents are *assumed* to know the full consequences of adopting each possible course of action or where economic agents exhibit `beliefs-rationality´. Therefore, whereas we argued in the previous chapter that Savage´s framework is an instance of *RPo*, we believe instead that Ramsey´s approach is an instance of *RPs.* Finally, last row in Table 2 below refers to the framework Keynes adopted in his later economic writings and which will be discussed in detail below. As some commentators have recognized, in his later economic writings Keynes adopted an approach where economic agents are viewed as falling back on social conventions that determine both expectations and confidence formation. Though Keynes does not refer explicitly in those writings to probability theory, it has been suggested by Gillies (2000) that he adopts an *inter-subjectivist* interpretation of probability insofar as probability is subjective (as in Ramsey´s theoretical framework) and the result of a complex process of interaction among the (interdependent) beliefs of a very large number of individuals. Finally, we argue in subsequent sections that the notion of rationality Keynes adopted in his later economic writings represents an instance of Popper´s *RPs*.

4. Keynes after Ramsey: the nature and role of conventions and the methodology of economics

We showed above that Keynes´s explicit capitulation to Ramsey´s views in *My Early Beliefs* (Keynes, 1973*a*, pp. 433-51) heralded a new stage in his philosophical development. However, Keynes does not provide an explicit philosophical account of his views on human nature (Davis, 1994, p. 147). Since economics was the main focus of his intellectual activity in the last part of his life it is to economics that one must turn to investigate his later views on human nature. In this respect, it is widely recognized that, in his later economic writings, Keynes adopted a theory of conventional judgement and belief.[26] We have argued above that the former exhibits a remarkable similarity with Hume´s dictum that human behaviour is ruled by `custom and habit´. Notwithstanding it, we will argue below that Keynes´s theory of conventional judgement and belief can be adjusted to make it compatible with PTKL. We also address below Keynes´s later ideas on both the methodology of economics and the nature of its object of study as they emerge from a number of letters addressed to Tinbergen, Kahn and, especially, Harrod in the late 1930s.

4.1. The nature and role of conventional knowledge

An extensive discussion of the philosophical foundations of Keynes´s thought is in Davis (1994). Along with several other commentators he argues that, in the wake of Ramsey´s critique of his approach to probability in the *Treatise*, Keynes concluded that individual judgement required an anchoring in `convention and traditional standards and inflexible rules of conduct´ and that the only resource upon which people might draw to correct `individual errors in judgement rested in a society´s accumulated conventions, standards, and rules´ (*op. cit.*, p. 104). Now, there are some characteristics of a society´s conventions, standards and rules referred to in Keynes´s later economic writings upon

which Davis (1994, ch. 4) focusses: (i) that they go beyond the simple summaries of past experience that the frequency approach to probability makes central to probability, (ii) that they possess a *normative* character above their quality as statistical description, (iii) that they possess *objectivity* as opposed to Ramsey´s subjectivist view of individual judgement as `degree of belief´, (iv) that they stem from *interdependent* judgement or mass psychology and, therefore, they presume the existence of a mechanism whereby individuals revise their judgements when they come into contact with the judgement of others and, finally, (v) that the formation of conventional expectations and confidence is a *dynamic* process. Our discussion below will focus on features (iii) through (v).

Now, according to Crotty (1994, p. 119), Keynes provides two main reasons for people´s observed reliance on conventions, standards and rules. The first reason is that he views agents as 'socially and endogenously-constituted human beings' whose actual behaviour changes with circumstances in a *dialectical* and *interactive* way. The second reason is that conventions *lull our disquietude* by enabling us `to hide from ourselves how little we can foresee´. In other words, the presence of radical uncertainty provokes disabling anxiety and the adoption of conventions helps mitigating this consciousness. As psychologists would put it, we do have a need to reduce our 'cognitive dissonance' (Festinger, 1957) and, to calm our nerves down, we collectively develop a conventional process of expectations and confidence formation. In the terminology of Choi (1993), the difficulty of finding an adequate 'paradigm' for ourselves encourages us to observe the paradigms that *other people* appear to use.[27] However, as Winslow (1989, p. 1180) notes 'conventions do not provide a rational foundation for long-run expectations; they are just "market place idols"' which help us substituting temporarily for the absence of knowledge. According to Keynes (1936, p. 152), the chief convention is the assumption that `the existing state of affairs will continue indefinitely, except in so far as we have

specific reasons to expect a change´ (Keynes, 1936, p. 152). However, the most detailed discussion of conventional judgement and belief is in Keynes´s 1937 *Quarterly Journal of Economics* paper where he delineates the ways in which individuals `save their faces´ as rational agents: [28]

1. We assume that the present is a much more serviceable guide to the future than a candid examination of past experience would show it to have been hitherto. In other words, we largely ignore the prospect of future changes about the actual character of which we know nothing.

2. We assume that the existing state of opinion as expressed both in prices and the character of current output is based on a *correct* summing up of future prospects, so we can accept it as such unless and until something new and relevant comes into the picture.

3. Knowing that our own individual judgement is worthless, we endeavour to fall back on the judgement of the rest of the world which is perhaps better informed. That is, we seek to conform to the behaviour of the majority or the average. The psychology of a society made up of individuals each of whom is attempting to copy the others leads to what we may term a *conventional* judgement (Keynes, 1937, pp. 214-15).

Now, a few comments are in order. First, the assumption that agents believe that `the existing state of affairs will continue indefinitely, except in so far as they have a specific reason to expect a change´ captures, synthetically, bullet points 1 and 2 in the

previous quotation. Yet, if we interpret this assumption as a behavioural hypothesis, it will only have empirical content if it is supplemented by a further assumption that tells us *how often* people have a specific reason to expect a change. For instance, if we argue that agents *almost never* have a specific reason to expect a change then, to the extent that individual behaviour tends to exhibit a stable pattern, the behavioural assumption will have high empirical content in the sense of exhibiting a high degree of testability. By contrast, if we argue that agents have a specific reason to expect a change *very often* so that individual is unlikely to follow a stable pattern, then the behavioural assumption has low empirical content (i.e., it exhibits a low degree of testability) and, hence, will be of little use for modelling purposes. Presumably, Keynes (1937) implicitly assumed that economic agents *normally* do not have a specific reason to expect a change so that, most of the time, they tend to assume that the `future will resemble the past´. Be that as it may, the frequency with which agents relax this assumption in the light of events that make them change their mind and believe that `the future is unlikely to resemble the past´ is an empirical issue and, to the best of our knowledge, Keynes does not express his views on it.

Our second comment is related to the `homogeneity of human behaviour through time´ implied by bullet points 1 and 2 above. As we will see in the following section, the main point of Keynes´s epistolary exchange with both Tinbergen and Harrod is his argument that the *material of economics is non-homogeneous through time* which then becomes the core thesis of his attack on the methodological stance adopted by Robbins and which Keynes identifies with the `natural science view´ (Keynes, 1973*b*, pp. 299-300). The alleged non-homogeneity through time of the material of the social sciences stems from the fact that `knowing that our own individual judgement is worthless, we tend to fall back on the judgement of the rest of the world´ and that the `psychology of a

society made up of individuals each of whom seeks to copy the others´ tends to make their judgement change unpredictably over time (Keynes, 1973*b*, pp. 214-15). In turn, the conventional nature of beliefs makes expectations and confidence exhibit instability as in a bootstrap equilibrium. That said, a commentator might object that Keynes may have incurred in a logical contradiction here. To be sure, if economic agents *normally* believe that `the existing state of affairs will continue indefinitely´ this will impart the economy with a high degree of stability and, hence, one would expect that the economic material exhibits a *high* (rather than a low) degree of homogeneity through time. However, we will argue below that this contradiction is only apparent and can be duly addressed if it can be shown that Keynes´s theory is an instance of *RPs*.

Our third comment is related to our previous argument above that the approach to human nature that Keynes adopts in his later economic writings is similar to Hume´s in the sense that, despite the absence of an `inductive logic´, Keynes, like Hume does, assumes that people tend to believe that the `future will resemble the past´. To be sure, bullet points 1 and 2 above capture the idea Keynes *assumes* that economic agents tend to extrapolate the present into the future unless they think they have a specific reason for not doing so. However, it is the notion of *interdependent* beliefs (bullet point 3) that draws special attention in Keynes´s later economic thought. As Davis (1994, p. 145) writes, `what is relevant here from that discussion [on the role of conventions] is the idea that conventions play a role alternative to inductive methods in enabling us to account for how the future is based on the past´. Rather, within the context of Keynes´s theory of conventional judgement, induction must be viewed as `being relative to the reasoning particular individuals exercised in particular historical contexts´ (*op. cit.*, p. 146). Keynes´s ideas into the conventional ways in which individual beliefs are formed are shown in the passage of the *General Theory* where he displays his `Beauty Contest´

metaphor and in which he explains the complex interactive mechanism through which the prices of financial assets are determined:

`Professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view… It is not a case of choosing those which, to the best of one´s judgement, are really the prettiest, or even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees´ (Keynes, 1936, p. 156).

As Davis (1997, p. 209) remarks, `the idea that it is not the prettiest but those most likely to be thought the prettiest nicely captures the change in Keynes´s view of judgement´.[29] Unlike the notion of `rational degree of belief´ he proposes in the *Treatise on Probability*, now the individual no longer intuits the [Platonic] quality of beauty, but rather she endeavours to form an interdependent judgement about what others facing the same dilemma may choose. In particular, each participant in the `beauty contest´ makes use of *introspection* to consider her own opinion, compares it to an opinion imputed to others, makes the necessary adjustments, and finally comes up with a new individual judgement that bears the imprint of mass psychology. In other words, according to this

view, individual beliefs are constructed within a complex structure of *interdependent* beliefs that works through an inter-subjective learning process whereby each individual 'puts herself in the shoes of others' to discover others´ beliefs.

4.2. The `rationality´ of conventional knowledge

The notion that individual behaviour tends to conform to the behaviour of the majority suggests that the former is based, at least partly, upon the *imitation* of others.[30] However, `rational´ economic agents will only tend to imitate behaviour they perceive as having proved *successful* in the past when was adopted by others. In turn, the former will normally be behaviour grounded on beliefs which have withstood the operation of trial and error-elimination processes. Keynes does not explain how some conventions come to be replaced by others and, hence, what we say here has to be understood as one possible (and certainly not the only one) development of Keynes´s ideas on this issue.[31] Be that as it may, we believe that this characterization of conventional behaviour under uncertainty is compatible and, indeed, *complementary* with PTKL.[32] As we explained in the previous chapter, the notion of rationality that stems from PTKL is characterised by *corrigibility*, i.e., the ability to eliminate one´s mistakes. Now, if individual beliefs are formed by comparing our beliefs to other people´s beliefs and, especially, to the beliefs of those people who appear to have been successful in particular endeavours in the past and, at the same time, those beliefs which are thought to have been discredited by the facts are not adopted, it turns out that the formation of conventional beliefs will tend to take place in a way that is reminiscent of PTKL.[33] Indeed, we believe that conventions, standards and rules can be viewed as *conjectures* or tentative hypotheses which emerge spontaneously as a response to a specific problem by virtue of a complex interactive process in which a large number of people participate. However, unlike the conjectures which are the usual

object of PTKL (for instance, in the case of theories of scientific progress), conventions, standards and rules exhibit a profound *social* dimension in the sense that they are the outcome of a complex process of interdependent beliefs among a large number of individuals.

This profound social dimension and the fact that *only* those conjectures which agents believe not to have been discredited in the past will be adopted suggests that the conjectures which play a key role in Keynesian macro-theory exhibit a peculiar nature. For lack of a better name, let us denote them as `hegemonic conjectures´ (HC hereafter). Like other conjectures, HC will be preserved as long as economic agents believe that they are useful for addressing certain problems — presumably because they see them as having been useful tools for guiding behaviour in the past — and, as long as they are preserved, they will confer stability and predictability to both individual and *collective* behaviour. In other words, we believe that the complex interactive process involving a large number of individuals that, arguably, characterises the formation of conventional beliefs in Keynes´s theory *is compatible with PTKL once the thesis that such interactive process is essentially based upon a trial and error-elimination mechanism is accepted*. In consequence, we think that *individual behaviour which broadly accords with HC as suggested above is `rational´ provided it exhibits corrigibility*. That is to say, according to PTKL, the charge of `irrationality´ only applies to those individuals who consciously refuse to eliminate their errors. Similarly, in the context of Keynesian theory, the charge of `irrationality´ *only* applies to those individuals who abide themselves by conventions, standards and rules which are widely known to have yielded undesirable results in the past.[34]

Our argument that Keynes´s theory of conventional beliefs as it appears in his later economic writings is, at least *a priori*, compatible with PTKL is not contradicted by the idea expressed by Davis (1997, p. 217) that, according to Keynes, `the conditions

associated with states of confidence concern the success or lack of success with which individuals come to assess each other´s opinions about markets´. That is, the success or lack of it is not related here to the relative past performance of behaviour based on the adoption of certain conventions but to the ability of individuals to reciprocally *discern* other people´s opinions. In other words, the idea we presented in the previous paragraph is that, if Keynes´s theory of conventional beliefs is to be made compatible with PTKL then this implies that people will tend to adopt those beliefs which appear not to have been discredited by the facts in the past whereas Davis (*op. cit*.) suggests that, according to Keynes, our *confidence* in a certain set of beliefs is a positive function of the success attained in reciprocally discerning other people´s beliefs. In other words, the degree of success would operate here at two different levels: beliefs (including expectations), and the confidence with which those beliefs are held on average. In the former case, it is the apparent degree of success exhibited by a set of beliefs in the past that will determine whether or not those beliefs become `hegemonic´. In the latter case, by contrast, it is the success in reciprocally discerning those beliefs, that is, the extent to which we manage to apprehend each other´s beliefs, that determines how confident we are about them. Be that as it may, we believe that the confidence with which people hold a set of beliefs will also depend positively on *how widely* that set of beliefs comes to be held across individuals. In turn, the *more widely held a set of beliefs is the more stable expectations will be* and, hence, it follows that the less widely held a certain set of beliefs is the more potential for macroeconomic instability there will be. With this, we believe we have a more encompassing theory of conventional judgement and belief with an evolutionary dimension and a capacity for enhancing and extending Keynes´s macro-theory. On the one hand, the notion that the likelihood that a set of beliefs comes to be widely adopted by economic agents depends positively upon the extent to which they are seen as not

having been discredited by the facts in the past provides Keynes´s macro-theory with an *evolutionary* theory of conventional belief anchored in PTKL. On the other hand, the notion that the (average) level of confidence with which a set of beliefs is held depends *positively* on the ability of economic agents to reciprocally discern each other´s beliefs and how widely those beliefs happen to be held provides Keynes´s macro-theory with an explanation for changes in the (average) level of confidence.[35]

4.3. Business cycles and the expansion of conventional knowledge over time

From our previous discussion it logically follows that HC may not be preserved indefinitely.[36] A HC will be replaced by another one whenever people *believe* that it no longer yields good results if adopted or when a sufficiently large amount of experience has accumulated which suggests that it is inadequate in the sense of apparently leading to undesirable results. If the convention to be replaced happens to be a convention upon which the evaluation of the long-term profit expectations of investors was based, then a period of crisis may ensue as investors increase substantially their *relative* demand for liquid assets (i.e., liquidity preference increases) in the aftermath of higher uncertainty about the future. In turn, an increase in the degree of liquidity preference will lead to a decrease in both the level of aggregate demand and economic activity. In the context of Keynes´s macro-theory, an increase in the degree of liquidity preference is concomitant to a decrease in the marginal efficiency of investment. The demand for investment will decrease as both the marginal efficiency of investment decreases and the term premium embedded in financial assets and bank loan rates increases. The decrease in aggregate investment will initially lead to a decrease in aggregate output and income and this, in turn, will cause a further decrease in aggregate spending as aggregate consumption goes down due to the initial fall in aggregate disposable income. This way, the breakdown of

a critical convention in the evaluation of long-term profits brings about a downturn in the economy. Finally, the turbulent period will come to an end when the old convention is replaced by a *new* one.[37]

4.4. The debate on the nature of economics and its object of study

A detailed discussion of Keynes´s involvement in several controversies related to the role of induction and statistical methods in economics is in Bateman (1990). The importance of this controversy goes beyond the role of statistical methods in economics insofar as it highlights the ideas that Keynes held on economic methodology over the latest part of his life. Of particular interest to us is the exchange that Keynes had with Tinbergen in the late 1930s because it took place when Keynes had already rejected most of the ideas he advocated in the *Treatise* and soon after he published the *General Theory* and his famous 1937 *QJE* paper. Keynes wrote a number of letters addressed to Tinbergen, Harrod, and Kahn where he expounds his ideas unambiguously. To this, we may add that the controversy is of high interest from the point of view of the history of economic thought since Keynes´s critique of Jan Tinbergen´s work represented the first systematic assessment of the multiple regression analysis the latter pioneered and which has had such a far-reaching impact on the ulterior development of economics.[38]

4.4.1. The non-homogeneity through time of economic data

It is not our purpose to reproduce in detail the terms of the exchange but only to summarize its main implications for Keynes´s views on methodology. As Bateman (*op. cit*.) notes, in the Tinbergen controversy Keynes argued that there is no *a priori* reason to expect that the economic data used in statistical analysis is stable in the long run and,

hence, that there is no reason for us to expect to infer stable correlations. This idea is expressed in the first letter Keynes addressed to the official of the League of Nations who had asked him to review Tinbergen´s work in August 1938:

`There is first of all the central question of methodology, —the logic of applying the method of multiple correlation to unanalysed economic material, which we know to be non-homogeneous over time. If we were dealing with the action of numerically measurable, independent forces, adequately analysed so that we knew we were dealing with independent atomic factors and between them completely comprehensive, acting with fluctuating relative strength on material constant and homogeneous through time, we might be able to use the method of multiple correlation with some confidence… In fact, we know that every one of these conditions is far from being satisfied by the economic material under investigation´ (Keynes, 1973*b*, pp. 285-86).

Further down in the same missive, Keynes adds:

`Is it assumed that the future is a *determinate* function of past statistics? What place is left for expectation and the state of confidence relating to the future?... What place is allowed for non-numerical factors, such as inventions, politics, labour troubles, wars, earthquakes, financial crisis? One feels a suspicion that the choice of factors is influenced (as is indeed only natural) by what statistics are available, and that many vital factors are ignored because they are statistically intractable or unprocurable´ (*op. cit*., p. 287, emphasis added).

Therefore, it seems that for Keynes, the alleged *non-homogeneity* of economic material through time stems from the fact that the latter can be affected by a potentially large array of *non-economic* factors which are not taken account of in statistical data. However, as we show below, Keynes also suggests in another letter addressed to Harrod in 1938 (Keynes, 1973*b*, p. 300) that such non-homogeneity of economic material over time stems from the fact that individual behaviour is based upon conventional beliefs which, by their nature, may change through time. In any case, Tinbergen´s answer was that it can be assumed that non-economic factors affect economic variables in a *non-systematic* fashion (i.e., they cancel each other out in the long run) so that their effect on the latter can be duly taken account of in the stochastic term of the statistical model. In any case, as Bateman (1990) notes, Keynes´s argument was not that the entire project of Tinbergen should be abandoned but that it needed to adopt a more inductive basis. In particular, as he explained twenty years before in Chapter 33 of the *Treatise*, he argued that, in order to carry out statistical inductions, we must previously examine series of series of economic data in order to determine whether the distributions of the variables are stable in as many circumstances as possible.[39] Be that as it may, Bateman (*op. cit*., p. 378) concludes that Keynes was not opposed to Tinbergen´s econometric work but only to attempts at statistical inference that were not preceded by a previous analysis aimed at ascertaining the suitability or otherwise of the data for making such inferences.

4.4.2. Keynes´s view of the nature of economics and its methodology

Next, after having shown above why Keynes thought that the economic material is not homogenous through time we turn to discussing his views on the methodology of economics. According to us, there are three main ideas that emerge from the two letters that Keynes addressed to Harrod in 1938 (Keynes, 1973*b*, pp. 295-300).[40] The first idea is his claim that `economics is a science of thinking in terms of models joined to the "art

of choosing models" which are relevant to the contemporary world´. The second idea is his claim that the objective of economics is not so much to `predict´ but to explain. The last idea is his notion of economics as a `moral science´ (in opposition to the `natural science view´ that Keynes attributes to Lionel Robbins) which deals with introspection, motives, expectations, and psychological uncertainties. We will argue below that these three ideas highlight that Keynes´s view of economic models exhibits some similarities with Popper´s.

Firstly, we noted in the previous chapter that, building upon Hayek´s distinction between `explanation in principle´ and `explanation in detail´ (Hayek, 1967*b*), Popper (1994, p. 163) distinguishes between the problems of explaining or predicting *singular events* and the problems of explaining or predicting a *kind* or *type* of event.[41] According to him, the former can be solved *without constructing a model* — only certain universal laws and the relevant initial conditions are needed — whereas the latter is most easily solved *by constructing a model* (*op. cit.*, p. 164). That is, Popper shares with Keynes the view that models are, arguably, the most important instrument for analysis in the social sciences, including economics. Furthermore, he argues that in the context of the social sciences models consists of `certain elements placed in a typical relationship to each other, plus certain universal laws of interaction — the "animating" laws´ (Popper, 1994, p. 165). It is at this point where there is a difference between Keynes and Popper in their respective views of models in that Popper is more specific than Keynes is as to *what* a model must contain: (i) certain elements of the P-S placed in a typical relationship to each other, and (ii) the `rationality principle´. By contrast, Keynes does not allude to an `animating law´ and says that `the object of a model is to segregate the semi-permanent or relatively constant factors from those which are transitory´. Yet, we will argue below that the `rationality principle´ is implicit in Keynes´s macro-theory.

Secondly, we have shown above that Popper argues that models seek to capture the *typical* aspects of a P-S with a view to making statements about a *type* of event and, consequently, that they represent something akin to *typical initial conditions* (*op. cit*., p. 164). He then adds that a statement about a *typical* event can be either an explanation of *why* that typical event occurred in the past, or a prediction. However, further down he makes it clear that `the fundamental problem of both the theoretical and the historical social sciences is *to explain and understand events in terms of human actions and* [typical] *social situations´* (Popper, *op. cit*., p. 166). The latter leads Notturno (1998, p. 412, emphasis added) to argue that the `problem of situational analysis in the theoretical and historical social sciences, in Popper´s view, is not to construct models that predict or prophesize the future; it is to construct models that help us *explain and understand the past´*. Now, we think we need to distinguish at this stage between `predictions´ and `forecasts´. As Beinhocker (2013) notes, in the theoretical social sciences, a `prediction´ (as opposed to a `forecast´) amounts to deriving the *deductive logical consequences* of a theory. When Popper makes a distinction between problems of explaining or predicting *singular events* from problems of explaining or predicting a *type* of event, we think he really means that forecasting, as opposed to deriving the deductive logical consequences of a theory, is *not* generally possible in the social sciences.[42] More specifically, the main problem of using models in the social sciences for forecasting purposes stems from the fact that they use *typical* initial conditions. Accordingly, we believe that Popper makes use of the term `prediction´ in the quotation above to refer to a legitimate objective of both the natural and the social sciences but his next comment further down in the same essay suggests he should have distinguished between `prediction´ in the natural sciences which amounts to `forecasting´ and `prediction´ in the social sciences which consists of the derivation of the deductive logical consequences of a theory. If our interpretation is

correct, then Popper´s position is rather similar to the position Keynes advocates when he makes it clear that `the object of statistical study is not so much to fill in missing variables with a view to prediction, as to test the relevance and validity of the model´ (Keynes, 1973*b*, pp. 295-6).

The third and last idea we identified above is Keynes´s notion of economics as a `moral science´ in contrast to the `natural science´ view that Keynes seems to attribute to Robbins. Firstly, Keynes argues that `the pseudo-analogy with the physical sciences leads directly counter to the *habit of mind* which is most important for an economist proper to acquire´ (Keynes, 1973*b*, p. 299, emphasis added). Next, he emphasizes that economics is a moral science´ and makes it clear that it deals with *introspection, values, motives, expectations and psychological uncertainties* (*op. cit*., p. 300). He then closes the paragraph with a metaphor whereby he explains why the material of economics is not homogeneous through time:

`One has to be constantly on guard against treating the material as constant and homogeneous. It is as though the fall of the apple to the ground depended on the apple´s motives, on whether it is worth while falling to the ground, and whether the ground wanted the apple to fall, and on mistaken calculations on the part of the apple as to how far it was from the centre of the earth´ (Keynes, 1973*b*, p. 300).

That is, he seems to argue that the non-homogeneity through time of the material of economics stems mainly from the fact that, being a `moral´ science, economics deals with the motives, expectations and psychological uncertainties of economic agents, all of which are *subjective* and, crucially, liable to experience unpredictable changes.[43] By

contrast, and building upon Hayek (1942, 1943), we argued in the previous chapter that the `natural´ sciences deal with *objective* facts in the sense that their nature is (largely) independent of the viewpoint of the observer and which, by their nature, are unlikely to undergo unpredictable changes. Now, we believe that the quotations reproduced above suggest that the `habit of mind´ that Keynes thinks is important for economists (and we add social scientists) to acquire is precisely the methodological position we associated in the previous chapter with *RPs,* i.e., with the view that the social scientist should try to reconstruct P-S not as she believes it is (objectively) but, instead, as she believes that agents *believe* it is.

5. Keynes´s views on economics and the `rationality principle´

The main purpose of this section is to argue that Keynes´s view of economics as a `moral´ science and the `natural science view´ he associates with Robbins (1932) exhibit an analogy with the `subjectivist´ and `objectivist´ version of *RP* respectively More specifically, we will argue below that Keynes´s macro-theory can be interpreted as an instance of *RPs*. We will also argue that Keynes´s attempt to distinguish his view of economics as a `moral´ science by emphasizing the role introspection plays in the former is misleading. In this respect, Robbins is known for his defence of the key role introspection plays in allowing us to grasp the premises upon which economic theory is grounded.[44] In other words, we think that introspection also plays a role in neoclassical economics and that the difference between Keynes´s view of economics as a `moral´ science and his view of neoclassical theory as an approach that employs a methodology that resembles the `natural´ sciences has to be ascribed to other characteristics of these approaches. In particular, we will argue below that the key difference between these two approaches lies in their respective adoption of *different* versions of *RP*.

5.1. The role of introspection in neoclassical economics

If we identify the approach Keynes associates with Robbins with neoclassical economics then, it is clear that, as Robbins (1932) argues, the role of introspection is to grasp the basic human dispositions upon which the theoretician must build her models. For instance, in terms of the framework presented in the previous chapter, introspection is the source of the notion that individual behaviour exhibits both `ends-rationality´ and `means-rationality´. In the case of `ends-rationality´, we noted above that, following von Mises (1944), human behaviour is *purposeful* or oriented towards the achievement of certain aims. Likewise, in the case of the notion of `means-rationality´, we defined it as the idea that individual behaviour is *consistent* for a given set of beliefs and objectives. We made it clear that fulfilment of ends- and means-rationality is common to models that adopt either *RPo* or *RPs* in that, in both cases, the behaviour of agents is assumed to be purposeful and adequate to the P-S. Yet, we also argued that a common feature of all the theoretical models that adopt *RPo* is that the theoretician `imposes´ her view of P-S upon the agents. This means that, in addition to fulfilment of both `ends-rationality´ and `means-rationality´, economic agents are also assumed to exhibit `beliefs-rationality´. In other words, agents are assumed to exhibit `substantive´ rationality (Simon, 1976). As we explained, the modern neoclassical version of this approach to rationality is captured in Subjective Expected Utility (SEU) theory (Savage, 1954), which entails that agents know the consequences of all possible courses of action under every `state of the world´ so that the only source of uncertainty is the (subjective) probability that agents attach to the latter. However, and unlike the notions of `ends-rationality´ and `means-rationality´, *we believe that the assumption of `beliefs-rationality´ does not stem from introspection. Rather, the latter is an `artefact´ of the theoretician aimed at closing her model in a way that testable predictions can be generated*. That is, by assuming that economic agents exhibit

`substantive´ rationality, the theoretician can `close´ her theoretical model in a way that clear testable predictions can be derived.[45] But, crucially, in order to do so, the theoretician has to ignore the fact that we can *never* be sure that our knowledge is `true´, i.e., that there is not an `inductive logic´. As we argue below, Keynesian theoreticians do not assume there is an `inductive logic´ and, for this reason, they do not assume that agents exhibit `beliefs-rationality´. We will also argue that, in order to circumvent this barrier they resort to a particular type of introspection; they `put themselves in the shoes of economic agents´ so that they can get an idea of *how* they make decisions in the face of uncertainty.

5.2. Keynesian uncertainty and the `logic of induction´

We have argued that the adoption of the assumption of `beliefs-rationality´ in neoclassical economics logically implies that the theoretician *assumes* that there is an `inductive logic´, i.e., she believes that she can obtain `true´ knowledge by extrapolating the information of the past into the future.[46] `True´ knowledge in this context means that every economic agent is (i) assumed to possess an exhaustive list of possible `states of the world´, and (ii) can attach numerical probabilities to each of them. But, surely, *such alleged `true´ knowledge can only come from the past.* Where else could it come from? By contrast, Hume (2006[1748]) showed long time ago that such an `inductive logic´ does not exist. Thus, in neoclassical models — or the models that Keynes associates to the `natural´ science view — the theoretician reconstructs P-S `as if´ agents believed that an `inductive logic´ exists.

We have noted that the assumptions of `ends-rationality´ and `means-rationality´ are shared by models that adopt *RPo* and *RPs* and, consequently, that all such models make use of introspection. The difference between neoclassical and Keynesian models

must thus lie elsewhere. Unlike neoclassical theoreticians, Keynesian theoreticians do not assume that agents exhibit `beliefs-rationality´. Rather, they assume that agents have to make decisions in a context of `uncertainty´ in the sense of both Keynes (1920) and Knight (1971[1921]).[47] That is, as opposed to SEU theory, in Keynes´s macro-theory economic agents do not possess an exhaustive list of possible `states of the world´, do not know the full consequences of adopting different courses of action under different circumstances and cannot attach numerical probabilities to all the possible `states of the world´. The reason for the existence of `uncertainty´ is usually associated to Keynes´s assumption that `economic data is not homogenous through time´ or, in more technical terms, to the non-ergodic nature of the world.[48] The latter implies that the `future will not resemble the past´ and, hence, the impossibility of knowing the full consequences of adopting different courses of actions under different `states of the world´. However, *we believe that the ultimate cause of the existence of genuine uncertainty in the economy is not so much the absence of homogeneity through time of economic data but the absence of an `inductive logic´*. In other words, `uncertainty´ is ultimately implied by Popper´s claim that all knowledge is *conjectural* and, hence, provisional. Even if we observe that economic data is homogeneous through a long period of time we cannot claim that the former will continue to be homogeneous in the future. In the absence of an `inductive logic´ we can never claim that `the future will resemble the past´. Such claims can only be *conjectural*. However, this may also be applied to the opposite claim, that is, to the claim that economic data is *not* homogeneous through time. To be sure, if Keynes´s claim that economic data is not homogeneous through were `true´ the former would be a powerful reason for making us be uncertain about the future. Yet, the former is only a conjecture that may eventually turn out to be false. That is, it is a conjecture. Thus, there is a *deeper* reason for being uncertain about the future than the possibility that the world

is not ergodic; *we cannot guarantee that the `future will resemble the past´*. In short, according to us, the primary source of genuine uncertainty in a market economy is not the alleged non-homogeneity through time of economic data but the fact that, due to the lack of an `inductive logic´, we cannot guarantee that the `future will resemble the past´.

5.3. Mainstream criticisms of Keynesian macro-theory

By assuming that the `future will not resemble the past´ Keynesian theoreticians have been exposed to the criticism that, since economic agents cannot confidently use the information obtained from the past to prophesize the future, their beliefs cannot be fully specified in their models (i.e., the theoretician cannot assume that economic agents exhibit `beliefs-rationality´) and, hence, no forecasts can be derived because economic agents´ decisions will depend on `exogenous´ and, hence, unpredictable changes in their beliefs. In particular, Keynesian macro-theory has often been denoted as `nihilistic´ by some neoclassical economists (Coddington, 1982) allegedly because, by assuming that the `future will not resemble the past´, Keynesian macro-models are unable to generate testable predictions. However, we believe the accusation of `nihilism´ is unjustified. As we explained above, the implicit adoption by neoclassical economists of a modelling approach based on *RPo* and the construction of models based upon the assumption that economic agents exhibit `substantive´ rationality allows them to generate neat testable predictions.[49] Yet, we argued in the previous chapter that the adoption of *RPo* implies that the theoretician `imposes´ upon the agents her view of P-S. More specifically, and given the absence of an `inductive logic´, we argued that the assumption that economic agents exhibit `beliefs-rationality´ entails the `imposition´ of the theoretician´s view of P-S upon them.

Next, Keynesian theoreticians *do adopt RPs instead of RPs*. As we argued in the previous chapter the adoption of *RPs* implies that the theoretician seeks to reconstruct P-S as `she believes that economic agents believe it is´. We have also argued above that Keynesian theoreticians *believe* that the `future will not resemble the past´. Rather, *they believe that economic agents tend to believe the opposite*. More specifically, Keynesian theoreticians observe that, in spite of the fact that the `future will not resemble the past´, this does not prevent economic agents from actually making decisions. According to them, economic agents *make decisions because, most of the time, they `ignore´ that the `future will not resemble the past´. Furthermore, they `ignore´ it for good reasons; if they didn´t, they would be paralysed!* At this stage, Keynesian theoreticians are `forced´ to rely on introspection in order to adopt agents´ viewpoint, i.e., they adopt *RPs*. That is, instead of `imposing´ upon agents their view of P-S — which as we argued above is the strategy implicitly adopted by neoclassical economists — they `put herself in the shoes of economic agents´ in order to understand *how* they make (often good) decisions in spite of the genuine uncertainty they face. For lack of a better name let us denote this type of introspection as *empathic* introspection.

We may add to this that, although the adoption by Keynesian theoreticians of the viewpoint of agents stems from their need to understand *how* the latter make decisions in spite of finding themselves in a situation of genuine uncertainty there is an even more fundamental reason for doing so. As we remarked in the previous chapter, Hayek (1942, 1943*a*) claims there are no objective `facts´ in the social sciences because, in the latter, `things are what people believe they are´.[50] However, if `things are what people believe they are´, it seems to us that the natural strategy for the theoretician is to `put herself in the shoes of agents´. It was for this reason that we argued in the previous chapter that the adoption of *RPs* is the natural approach in the social sciences. To this, let us add that

Hayek (1943*a*) insists that the theoretician does not possess herself any information that is not possessed by the agents themselves. Therefore, the adoption of *RPs* rather than *RPo* in the social sciences can be advocated not only on the grounds that, owing to the lack of an `inductive logic´, the theoretician needs to rely on `empathic´ introspection to acquire knowledge about *how* economic agents make decisions in a context of genuine uncertainty. The adoption of *RPs* instead of *RPo* can also be defended on the grounds that: (i) the P-S that agents face is `as they see it´ rather than as the theoretician sees it, and (ii) the theoretician does not possess any information that is not possessed by the agents.

Next, the assumption Keynes (1936, p.152) makes according to which economic agents believe `the existing state of affairs will continue indefinitely, except in so far as they have specific reasons to expect a change´ does not prevent him, however, from generating testable predictions. We thus believe that the charge of nihilism thrown upon Keynesian macro-theory by neoclassical economists is unwarranted. Some examples of testable predictions that stem from Keynesian macro-theory are the following: (i) the inability of money wage cuts to eliminate involuntary unemployment and their potential destabilizing effects, (ii) the ineffectiveness of monetary policy during recessions, and (iii) the existence of involuntary unemployment due to too high real (long-term) interest rates rather than to `too high´ real wages. These predictions have to be understood as the `logical implications´ of the theory which stem from the premise that economic agents´ long-term expectations and confidence are formed *conventionally* and, hence, they are given in the short term. As Crotty (1994) argues, it is the stability and predictability that social conventions confer to individual and collective economic behaviour that provides the behavioural foundations for Keynesian macro-modelling. Specifically, the adoption of social conventions by agents helps generate orderly and stable patterns of economic

behaviour across individuals thus opening up the possibility of doing macroeconomic analysis. He denotes the stability generated by decision-making under uncertainty based on conventional behaviour as `conditional´ stability and defines it as a situation where *behavioural equations are stable under conditions that hold most of the time* but which, at the same time, as in the celebrated Minsky´s (1982) `financial instability hypothesis´, creates the conditions for occasional bouts of macroeconomic instability.

Lastly, the accusation by some mainstream economists that Keynesian macro-models are made up of economic agents who behave `irrationally´ is also, according to us, a direct consequence of the fact that, unlike in neoclassical economics, in Keynesian macro-theory economic agents are no assumed to exhibit `substantive´ rationality. In other words, Keynesian economic agents are not optimizers and, from the standpoint of neoclassical theory, this is tantamount to `irrationality´ (Becker, 1962). However, we believe the association of `rational´ behaviour with optimizing behaviour is arbitrary. To be sure, this notion of rationality does not stem from philosophy of science but from a concrete tradition in the history of economic thought: neoclassical economics. As we have argued above, such notion of rationality *presupposes* that there is an `inductive logic´ theoreticians can take advantage of. In particular, if the existence of an `inductive logic´ is assumed, the theoretician can assume that she can acquire `true´ knowledge about P-S which, when combined with the knowledge she can obtain *via* introspection about the basic premises of human behaviour, allows her to construct models where she `imposes´ upon the agents her alleged `true´ view of P-S. However, if the existence of an `inductive logic´ is denied, such `true´ view of P-S becomes unattainable. We have argued above that Keynesian theory does not presume the existence of an `inductive logic´. In particular, we argued that Keynesian theoreticians believe that the `future will not resemble the past´ yet they realize that economic agents make decisions in spite of it. It

is for this very reason that Keynesian theoreticians adopt *RPs*, that is, they resort to `empathic´ introspection in order to understand *how* agents behave. Their main result is that economic agents fall back on social conventions to make decisions. But, crucially, reliance by economic agents on conventional knowledge in the absence of an `inductive logic´ does by no means imply that their behaviour is `irrational´. Rather, in the absence of an `inductive logic´ such behaviour is `rational´ insofar as the former is *coherent with economic agents´ belief in the presence of a high degree of stability and predictability in economic affairs*. We conclude that the charge of `irrational´ behaviour by economic agents in Keynesian macro-models made by some neoclassical economists is arbitrary and, hence, unwarranted.

5.4. Keynesian macro-theory in the light of Popper´s Rationality Principle

The modelling technique adopted by Keynesian theoreticians which we denoted above as `empathic´ introspection constitutes a heuristic device the theoretician relies on to gain understanding of what is a very complex process.[51] Thus, in Keynesian macro-theory the role of introspection is two-sided: (i) to gain insight into the motivations of individual behaviour in the fashion of neoclassical economists by taking advantage, as Hayek (1943*b*) would put it, of the existence of an *analogy* between the mind of the theoretician and the mind of the other human beings, and (ii) to `put oneself in the shoes of others´ to understand *how* economic agents behave *in different circumstances*, i.e., to adopt their viewpoint. Now, according to Keynesian macro-theory, there are two main *typical* situations economic agents may `believe´ to find themselves in:

1. Situations characterised by the absence of a significant discrepancy between expectations and realised outcomes which makes economic agents `believe´

that the past is a *good enough* guide to the future. In such situations, economic agents will tend, perhaps unconsciously, to extrapolate the past into the future. Confidence will be high and liquidity preference low. These are the so-called `tranquillity´ periods.

2. Situations characterised by the presence of a significant discrepancy between expectations and realised outcomes which makes economic agents `believe´ that the past *is not a valid guide to the future*. The former will typically lead to the breakdown of previously-held conventions as well as to the search for new conventions upon which to ground decision-making. Confidence will be low and liquidity preference high. These are the so-called `turbulence´ periods.

As we noted above, `tranquillity´ periods provide the foundations for Keynesian macro-modelling. In particular, and following Crotty (1994), we denoted the stability that results from decision-making under uncertainty based on conventional behaviour as `conditional´ stability. In turn, we have pointed out that `conditional´ stability refers to a situation where behavioural equations are stable under conditions *that hold most of the time but which may unpredictably `break down´* in the aftermath of the occurrence of occasional bouts of macroeconomic instability. Consequently, Keynesian macro-theory recognizes the inherent instability of decentralized market economies and, hence, the inability of macroeconomic models to become a tool for macroeconomic forecasting. The purpose of theory is not, according to Keynesian theory, to generate quantitative forecasts but to generate `predictions´ or logical implications of the theory which help us understand *how* market economies work and which can, in turn, serve as a basis for policy analysis.

Next, there are several elements of Keynes´s macro-theory which, according to us, lend support to our claim that it constitutes an instance of *RPs*. First, and foremost, there is his `Beauty Contest´ metaphor (Keynes, 1936, p. 156). The standard theory in economics for the formation of financial asset prices is that the latter represent the discounted stream of the future expected capital income they will generate for asset holders. But this is how the theorist *believes* financial asset prices are `objectively´ set. By contrast, as we have noted above, Keynes argues that the prices of financial assets are determined largely by the *speculative* behaviour of financial market participants. In particular, he claims that the decision of market participants to either sell or purchase a financial asset depends on: (i) her perception as to whether there is a `bull´ or a `bear´ market (i.e., whether she expects asset prices to rise or fall in the near future) and, (ii) particularly, her beliefs about how the other market participants think that asset prices are likely to behave in the near future as well as her beliefs about *how* other market participants believe that other market participants, in turn, expect asset prices to behave. To summarize, Keynes´s theory of financial asset prices is built upon the notion that the latter are set by a complex interactive process characterised by the interdependency of conventional beliefs among a very large number of financial market participants. But, in our view, this insight stems from the fact that Keynes a*dopts the point of view of market participants to understand how they behave* or, put another way, he `puts himself in the shoes of financial market participants´.[52] Yet, `putting himself in the shoes of economic agents´ is, according to us, what the macroeconomist does if he adopts *RPs*.

According to us, there are other instances of the theoretician `putting herself in the shoes of the agents´ in Keynes´s theory. Firstly, there are the so-called `fundamental psychological laws´ to which he alludes in his *General Theory*: (i) the psychological propensity to consume,[53] (ii) the psychological attitude to liquidity preference, and (iii)

the psychological expectation of future yield from productive capital-assets. The first law corresponds to households, the second one corresponds to speculators, and the third one corresponds to entrepreneurs. In all three cases, the behaviour of economic agents towards consumption, liquidity and long-term investment is a `rational´ response to the *uncertainty* they face. However, it is in the last two psychological laws that the adoption by the theoretician of the viewpoint of economic agents is clear. To be sure, economic agents´ liquidity preference and long-term profit expectations are closely related to their `confidence´ where the latter is a *subjective* measure of the `degree of completeness´ of the factual evidence on which their (uncertain) beliefs are based.[54] Specifically, the lower agents´ confidence is, the lower their demand for investment and the higher their liquidity preference will be. Furthermore, such behaviour is `rational´ in the presence of uncertainty.[55] To be sure, the notion that a fall in the (subjective) `state of confidence´ of economic agents is coupled to an increase in liquidity preference and a decrease in investment demand reflects the (implicit) decision of the theoretician to adopt agents´ viewpoint. Secondly, there is Keynes´s argument for the downward rigidity of money wages which he ascribes to the fact that workers care, mostly, about *relative* (money) wages so that they will not oppose an across-the-board decrease in *real* wages brought about by a rise in the general price level yet they will ferociously oppose attempts to cut their *money* wages down since the latter will presumably occur in a piecemeal fashion, i.e., money wage cuts will not take place at the same time but will occur sequentially. Lastly, Keynes´s important distinction between short-term and long-term expectations is, arguably, another instance of what we have denoted as `empathic´ introspection. In this case, Keynes associates short-term expectations with the level of demand for their products that entrepreneurs expect to face in the short-term, where the latter denotes the time span necessary to complete the production process. As some scholars have noted, it

appears that Keynes implicitly accepted that entrepreneurs rely on simple extrapolative methods to estimate the demand for their output in the short run and that, consequently, the short-run production decisions of business firms are, in his macro-model, assumed to be endogenously-determined.[56] By contrast, he argues that long-term expectations depend on a large array of factors and, hence, he treats them as an `exogenous´ variable. Be that as it may, the key point is that Keynes´s distinction between short- and long-term expectations reflects *his adopting the viewpoint of entrepreneurs* when having to make production and investment decisions respectively.

6. Summary and conclusions

We showed in the previous chapter that Hume was sceptical about the rationality of human behaviour because he believed that the latter was grounded upon `custom and habit´, that is, on the assumption that the `future will resemble the past´ in spite of the inexistence of an `inductive logic´. This paradoxical situation was denoted by Hume as the `problem of induction´. We showed the `solution´ proposed by Popper to the latter which we denoted as Popper´s evolutionary theory of knowledge and learning (PTKL). Nevertheless, there is a prior-to-Popper attempt to `solve´ the `problem of induction´: Keynes´s (failed) attempt to provide an objective epistemological theory of probability in his *Treatise on Probability*. Keynes´s aim was not only to answer Hume but also to answer Cambridge philosopher Moore (1993[1903]) who had argued that the highest expected `good´ would result if individual behaviour consists of following rules insofar as the latter represent the accumulated knowledge in a society. Keynes´s proposal to solve the `problem of induction´ consists of replacing the notion of `certain´ knowledge, which Hume showed to be unattainable, by the notion of `probabilistic´ knowledge. Now, his argument was that, though we cannot attain the former, we can nevertheless attain the

latter. Keynes´s main innovation in the *Treatise* was the notion of `logical probability relations´. As Andrews (1999) has argued, the latter resemble Platonic Universals in the sense that they exist in a `logical space´, are independent of individual opinions, and can only be discovered *via* intuition. Yet, the existence of `logical probability relations´ was successfully criticized by Ramsey (1978[1931]) and Keynes capitulated to Ramsey´s critique in 1938 (Keynes, 1973*a*, p. 445) so that by the second half of the 1930s Keynes abandoned most of the ideas he had advocated in the *Treatise* and adopted a pragmatic approach to human behaviour which is remarkably similar to Hume´s views. In other words, he implicitly admitted the impossibility of providing a solution to the `problem of induction´. We showed that, in the same essay where he criticizes Keynes´s *Treatise*, Ramsey (*op. cit*.) sketches a theoretical framework which is believed to have served as a benchmark for Savage´s (1954) Subjective Expected Utility (SEU) theory. However, we argued that, whereas Savage´s framework adopts the `objectivist´ version of Popper´s *RP*, Ramsey adopts the `subjectivist´ version instead. Nevertheless, it does not seem that the framework sketched in Ramsey (*op. cit*.) inspired in any significant way Keynes´s later economic writings even though it is clear he read Ramsey´s essay and accepted his criticism of the *Treatise*.

Next, the view Keynes adopts in his later economic writings appears to consist of the notion that we make decisions by using inductive procedures which essentially amount to adopting the social rules and conventions which are widely believed to have yielded good results in the past. The rules and conventions Keynes identifies in his later economic writings are, for our purposes, equivalent to Hume´s old idea that individual behaviour is grounded upon `custom and habit´. In particular, the chief convention of all, according to Keynes (1937), is the assumption that `the existing state of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change´

(Keynes, 1936, p. 152). We have argued that this assumption plays an essential role in Keynesian macro-theory in that it enables the theoretician to construct models in which: (i) economic agents behave in a `rational´ way in a context of genuine uncertainty, and (ii) testable predictions can *a priori* be generated.

The main purpose of this essay was to show that Keynesian macro-theory can be provided with strong epistemological foundations. In particular, we have argued that the former is essentially compatible with PTKL. Furthermore, we have also argued that the methodology underlying Keynes´s macro-theory can be interpreted as an instance of the `subjectivist´ version of Popper´s `Rationality Principle´ (*RPs*). As to the former case, we tried to show that the formation of both conventional expectations and confidence in Keynes´s macro-theory can be viewed as the ultimate result of a complex social process characterized by trial and error-elimination of beliefs based on conventional knowledge where those conventions that are believed by economic agents to lead to the making of wrong decisions tend to be replaced by new conventions that emerge spontaneously. In this respect, we coined the term `hegemonic conjectures´ to denote those conventions that are used by economic agents because they believe them to represent a valid guide for decision-making. The term `hegemonic conjectures´ captures two different ideas: (i) that the rules and conventions economic agents rely on resemble tentative hypotheses or conjectures in the sense that they are *provisional* and, hence, do not represent a claim to be `true´ knowledge and, (ii) that they become `hegemonic´ or widely used by economic agents through a complex social process. To the extent that the process whereby these social conventions emerge and die out is based on trial and error-elimination, we argued that the growth of conventional knowledge over time in Keynes´s theory resembles the evolutionary process of expansion of knowledge in general in PTKL. In this respect, we also suggested that the evolution of conventional knowledge in Keynes´s macro-theory is

analogous to Kuhn´s (1962) theory of scientific evolution insofar as the mechanism whereby a certain set of conventions is replaced by another one in the former exhibits a number of similarities with the mechanism whereby a scientific paradigm is superseded by another one in the latter. In particular, we believe that the role played by `anomalies´ in Kuhn´s framework is similar to the role played by `confidence´ crisis in Keynes´s macro-theory. As we explained, a `confidence´ crisis can be seen as taking place in the wake of the occurrence of a large observed discrepancy between expected and realized outcomes so that the social conventions that agents have been using for some time are no longer widely believed to represent a valid guide for decision-making. However, we also noted that there may be some conventions which remain `hegemonic´ indefinitely because of their simplicity and proved effectiveness. For instance, the three conventions Keynes refers to in his 1937 *Quarterly Journal of Economics* paper (Keynes, 1937) and which were the object of our discussion above are clear examples of social conventions which, to the extent that they remain in use indefinitely, provide the basis for Keynesian macro-modelling.

Next, we argued above that the convention that `the existing state of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change´ is at the core of Keynesian macro-theory and is directly responsible for two criticisms the latter has frequently been subject to: (i) the alleged `irrational´ behaviour of Keynesian economic agents, and (ii) the *nihilism* of Keynesian macro-theory. Notwithstanding it, we argued that these criticisms stem from the adoption by mainstream economists of the `objectivist´ version of Popper´s `Rationality Principle´ (*RPo*). Now, we showed in the previous chapter that, in the context of neoclassical economics, the former implies that economic agents are assumed to exhibit `substantive´ rationality (i.e., to be optimizers). Furthermore, the adoption of *RPo* implies that economic agents´ view of the `problem-

situation´ (P-S) is assumed to *coincide* with the theoretician´s. This, we argued, allows neoclassical economists to `close´ the model in a way that neat testable predictions can be derived. Yet, we argued that the `substantive´ rationality assumption presupposes that an `inductive logic´ exists that allows economic agents to acquire `true´ knowledge. Crucially, the adoption of *RPo* by neoclassical economists implies, according to us, that they assess Keynes´s macro-theory under the implicit assumption that there must be a coincidence between the agents´ and the theoreticians´ view of P-S. Thus, for instance, if the theoretician does *not* really believe that `the future will resemble the past´ — as in Keynes´s case — it follows that the construction of models where economic agents are assumed to *believe*, for practical purposes, that `the future will resemble the past´ is incoherent from the viewpoint of neoclassical theorists. Further, the charge of `nihilism´ by the latter stems from the fact that the possibility of generating testable predictions in Keynes´s theory is, according to them, precluded by the fact that agents´ expectations are assumed to be exogenous and, hence, to change unpredictably.

Now, we argued that the two charges alluded to above are unjustified and can be answered by showing that Keynes´s theory is, unlike neoclassical theory, an instance of *RPs*. As we explained in the previous chapter, the latter implies that the theoretician seeks to reconstruct P-S not `as she believes it is´ but, rather, `as she believes that gents believe it is´. First, the charge that Keynesian models posit economic agents who exhibit `irrational´ is related to the fact that, in Keynes´s macro-theory, agents are not assumed to exhibit `beliefs-rationality´. Yet, we argued that this is because, unlike neoclassical economists, Keynesian theoreticians (rightly) assume that there is no `inductive logic´ (i.e., `true´ knowledge cannot be attained). Besides, the adoption of *RPs* in Keynesian macro-theory implies, according to this interpretation, that though the theoretician does *not* believe that `the future will resemble the past´, she may nevertheless assume that

agents believe this is the best working assumption they can make owing to the presence of genuine uncertainty in the economy. It follows that the interpretation of Keynesian macro-theory we propound implies that it is not contradictory to assume that economic agents believe that `the existing state of affairs will continue indefinitely, except in so far as they have specific reasons for expecting a change´ even though the theoretician does *not* believe so. Likewise, if we assume that economic agents believe, for practical purposes, that the `future will resemble the past´, then Keynesian theoreticians cannot be accused of portraying economic agents´ behaviour as being `irrational´ since such behaviour is *coherent* with agents´ belief in the presence of a high degree of stability in the economy. Next, and according to our interpretation, Keynesian macro-theory cannot be accused of `nihilism´ either since the assumption that economic agents believe that `the existing state of affairs will continue indefinitely, except insofar as they have specific reasons to expect a change´ allows Keynesian theoreticians to construct macro-models that generate testable *predictions* (as opposed to quantitative forecasts). Some examples of these predictions as well as some examples of different parts of Keynesian macro-theory that illustrate the implicit adoption of *RPs* were provided. In short, our interpretation of Keynesian macro-theory as an instance of *RPs* is conducive to both a `rationalization´ of the typical criticisms the former is usually subject to as well as to an answer of them.

Finally, Keynesian economists tend to associate the existence of `uncertainty´ in the economy to the non-ergodic nature of the economic world. To be sure, the alleged non-ergodicity of the world is consistent with Keynes´s claim — which was the basis for his criticism of Tinbergen´s pioneering work on statistical inference in economics — that `economic data is not homogeneous over time´. By contrast, we argued above that a logical implication of our interpretation of Keynesian macro-theory is that the primary

cause for the existence of genuine uncertainty in the economy and its impact on agents´ behaviour is not so much the absence of homogeneity through time of economic data and, hence, the alleged impossibility of extrapolating the past into the future, but *the absence of an `inductive logic´*. In other words, `uncertainty´ is ultimately implied by Popper´s claim that all knowledge is *conjectural* and, hence, provisional. Consequently, even if economic data appears to be homogeneous through a long period of time, this is by no means a proof that the former will remain homogeneous in the future. That is, in the absence of an `inductive logic´ it cannot be shown that `the future will resemble the past´. Such claim is *conjectural*. This suggests that the relevance of the notion of non-ergodicity for Keynesian macro-theory is, according to us, overemphasised insofar as an even more substantial reason for the existence of uncertainty is the *conjectural* nature of human knowledge.

References

Andrews, D. R. (1999) Continuity and change in Keynes´s thought: the importance of Hume, *European Journal of the History of Economic Thought*, 6(1), pp. 1-21.

Arrow, K.J. (1986) Rationality of Self and Others in an Economic System, *Journal of Business*, 59(4), pp. S385-99.

Bateman, B. W. (1987) Keynes´s Changing Conception of Probability, *Economics & Philosophy*, 3(1), pp. 97-120.

Bateman, B. W. (1988) G. E. Moore & J. M. Keynes: A Missing Chapter in the History of the Expected Utility Model, *American Economic Review*, 78(5), pp. 1098-1106.

Bateman, B. W. (1990) Keynes, induction, & econometrics, *History of Political Economy*, 22(2), pp. 359-379.

Bateman, B. W. (1991) Das Maynard Keynes Problem, *Cambridge Journal of Economics*, 1(15), pp. 101-111.

Becker, G. (1962) Irrational behavior and economic theory, *Journal of Political Economy*, 70(1), pp. 1-13.

Beinhocker, E. D. (2013) Reflexivity, complexity, and the nature of social science, *Journal of Economic Methodology*, 20(4), pp. 330-342.

Brown-Collier, E. & Bausor, R. (1988) The Epistemological Foundations of the General Theory, *Scottish Journal of Political Economy*, 35(3), pp. 227-241.

Carabelli, A. M. (1988) *On Keynes´s Method*, London: Macmillan.

Carnap, R. (1950) *Logical Foundations of Probability*, Chicago: University of Chicago Press.

Carvalho, F. (1988) Keynes on probability, uncertainty, and decision making, *Journal of Post Keynesian Economics*, 11(1), pp. 66-80.

Choi, Y. B. (1993) *Paradigms and Conventions. Uncertainty, Decision Making, and Entrepreneurship*, Michigan: University of Michigan.

Coddington, A. (1976) Keynesian Economics: The Search for First Principles, *Journal of Economic Literature*, 14(4), pp. 1258-1273.

Coddington, A. (1982) Deficient Foresight: A Troublesome Theme in Keynesian Economics, *American Economic Review*, 72(3), pp. 480-487.

Cottrell, A. (1993) Keynes´s Theory of Probability and its Relevance to his Economics. Three Theses, *Economics & Philosophy*, 9(1), pp. 25-51.

Crotty, J. R. (1980) Post-Keynesian Economic Theory: An Overview and Evaluation, *American Economic Review*, Papers and Proceedings, 70(2), pp. 20-25.

Crotty, J. R. (1990) Keynes on the Stages of Development of the Capitalist Economy: The Institutional Foundations of Keynes´s Methodology, *Journal of Economic Issues*, 24(3), pp. 761-780.

Crotty, J. R. (1994) Are Keynesian Uncertainty and Macrotheory Compatible? Conventional Decision Making, Institutional Structures, and Conditional Stability in Keynesian Macromodels, in G. Dymski & R. Pollin (eds.), *New Perspectives in Monetary Macroeconomics*, pp. 105-139, Ann Arbor: University of Michigan Press.

Davidson, P. (1991) Is Probability Theory Relevant for Uncertainty? A Post Keynesian Perspective, *Journal of Economic Perspectives*, 5(1), pp. 129-143.

Davis, J. B. (1989-90) Keynes & Organicism: Comment, *Journal of Post Keynesian Economics*, 12(2), pp. 308-315.

Davis, J. B. (1994) *Keynes´s Philosophical Development*, Cambridge: Cambridge University Press.

Davis, J. B. (1997) J. M. Keynes on History and Convention, in Harcourt, G. C. and Riach, P. A. (eds.) *A 'Second Edition' of The General Theory*, Vol. 2, pp. 203-221, London: Routledge.

De Finetti, B. (1937) Foresight: Its Logical Laws, Its subjective Sources, *Annales de l´Institut Henri Poincaré*, vol. 7, reproduced in Kyburg, H. E. & Smokler, H. E. (eds.) *Studies in Subjective Probability*, pp. 93-158, 1967, New York: John Willey & Sons.

Dow, S. C. (1995) The appeal of neoclassical economics: some insights from Keynes's epistemology, *Cambridge Journal of Economics*, 19(6), pp. 715-733.

Dow, S. C. (2009) David Hume & Modern Economics, *Capitalism & Society*, 4(1), Article 1.

Dow, A. & Dow, S. C. (2011) Animal Spirits Revisited, *Capitalism & Society*, 6(2), Article 1.

Eichner, A. S. & Kregel, J. A. (1975) An Essay on Post-Keynesian Theory: A New Paradigm in Economics, *Journal of Economic Literature*, 13(4), pp. 1293-1314.

Festinger, L. (1957) *A theory of cognitive dissonance*, Standford, California: University of Standford.

Fitzgibbon, A. (1988) *Keynes's vision: a new political economy*, Oxford (UK): Clarendon Press.

Gillies, D. (2000) *Philosophical Theories of Probability*, Routledge: London.

Hamlin, A. P (1986) *Ethics, Economics and the State*, New York: St. Martin´s Press. Reprinted in Caldwell, B. (ed.), 1993, *The Philosophy and Methodology of Economics I*, pp. 304-362, Aldershot: Edward Elgar.

Hamouda, O. F. & Harcourt, G. C. (1988) Post Keynesianism: From Criticism to Coherence, *Bulletin of Economic Research*, 40(1), pp. 1-33.

Hamouda, O. & Smithin, J. (1988) Some Remarks on 'Uncertainty and Economic Analysis', *Economic Journal*, 98(389), March, pp. 159-164.

Hayek, F. A. (1942) Scientism and the Study of Society. Part I, *Economica*, 9(35), pp. 267-291.

Hayek, F. A. (1943*a*) The Facts of the Social Sciences, *Ethics*, 54(1), pp. 1-13.

Hayek, F. A. (1943*b*) Scientism and the Study of Society. Part II, *Economica*, 10 (37), pp. 34-63.

Hayek, F. A. (1967) Degrees of Explanation, in *Studies in Philosophy, Politics & Economics*, pp. 3-21, Chicago: University of Chicago Press.

Heiner, R. A. (1983) The Origin of Predictable Behavior, *American Economic Review*, 73(4), pp. 560-595.

Hempel, G. C. (1949) The Function of General Laws in History, in H. Feigl & W. Sellars (eds.) *Readings in Philosophical Analysis*, pp. 462-5, New York: Appleton-Century-Crofts .

Hicks, J. R. (1974) *The Crisis in Keynesian Economics*, Oxford: Basil Blackwell.

Hicks, J. R. (1979) *Causality in Economics*, New York: Basic Books.

Howitt, P. (1997) Expectations and Uncertainty in Contemporary Keynesian Models, in Harcourt, G. C. & Riach, P. A. (eds.) *A 'Second Edition' of The General Theory*, Vol. 1, pp. 238-260, London: Routledge.

Hume, D. (1978 [1739-40]) *A Treatise of Human Nature*, Oxford: Clarendon Press.

Hume, D. (2006[1748]) *An enquiry concerning human understanding: a critical edition*, Oxford, Clarendon Press; New York: Oxford University Press.

Iwai, K. (2009) The Second End of Laissez Faire ― Bootstrapping Nature of Money and Inherent Nature of Capitalism, *CIRJE-F-646 Discussion Paper*, August, University of Tokyo.

Jeffreys, H. (1939) *Theory of Probability*, Oxford: Oxford University Press.

Kakarot-Handtke, E. (2012) Why Post Keynesianism is not yet a science, Working Paper, *University of Stuttgart, Institute of Economics and Law*, September.

Keynes, J. M. (1920) *A Treatise on Probability*, London: Macmillan.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. Cambridge: Macmillan.

Keynes, J. M. (1937) The General Theory of Employment, *Quarterly Journal of Economics*, 51(2), pp. 209–223.

Keynes, J. M. (1973*a*) Essays in Biography, *The Collected Writings of John Maynard Keynes*, vol. X, London: Macmillan.

Keynes, J. M. (1973*b*) The General Theory and After: A Supplement, *The Collected Writings of John Maynard Keynes*, vol. XXIX, pp. 293-94, London: Macmillan.

Klant, J. (1985) The Slippery Transition, in T. Lawson & H. Pesaran (eds.) *Keynes´s Economics*, London: Croon Helm.

Knight, F. H. (1971[1921]) *Risk, Uncertainty, and Profit*, Chicago: The University of Chicago Press.

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago: The University of Chicago Press.

Lavoie, M. (2014) *Post-Keynesian economics: new foundations*, Northampton (MA): Edward Elgar.

Lawson, T. (1985) Uncertainty and Economic Analysis, *Economic Journal*, 95(380), pp. 909-927.

Lawson, T. (1987) The Relative/Absolute Nature of Knowledge and Economic Analysis, *Economic Journal*, 97(388), pp. 951-970.

Lawson, T. (1988) Probability and uncertainty in economic analysis, *Journal of Post Keynesian Economics*, 11(1), pp. 38-65.

Littleboy, B. (1990) *On Interpreting Keynes: A Study in Reconciliation*, London: Routledge.

Moore, G. E. (1912) *Ethics*, New York: Henry Holt.

Moore, G.E (1993[1903]), *Principia Ethica*, Cambridge: Cambridge University Press.

Notturno, M. A. (1998) Truth, Rationality, and the Situation, *Philosophy of the Social Sciences*, 28(3), pp. 400-421.

O'Donnell, R. (1990) An overview of probability, expectations, uncertainty and rationality in Keynes's conceptual framework, *Review of Political Economy*, 2(3), pp. 253-66.

Papandreou, A, G. (1959) Explanation and Prediction in Economics, *Science*, 129, April, pp. 1096-1100.

Popper, K. R. (1959) The Propensity Interpretation of Probability, *British Journal for the Philosophy of Science*, 10 (37), pp. 25-42.

Popper, K. R. (1990) *A World of Propensities*, Bristol (UK): Thoemmes.

Popper, K. R. (1994) *The Myth of the Framework: In defence of science and rationality*, London: Routledge.

Ramsey, F. P. (1978[1931]) 'Truth and probability', in *Foundations: Essays in Philosophy, Logic, Mathematics & Economics*, pp. 58-100, edited by D. H. Mellor, London: Routledge & Kegan Paul.

Robbins, L. (1932) *An Essay on the Nature and Significance of Economic Science*, London: Macmillan. Reprinted in D. M. Hausman (ed.) 1984, *The Philosophy of Economics: An anthology*, pp. 113-140, Cambridge: Cambridge University Press.

Rotheim, R. J. (1989-90) Organicism and the role of the individual in Keynes: Comment, *Journal of Post Keynesian Economics*, 12(2), pp. 316-326

Runde, J. (1990) Keynesian Uncertainty and the Weight of Arguments, *Economics and Philosophy*, 6(2), pp. 275-292.

Runde, J. (1997) Keynesian Methodology, in Harcourt, G. C. & Riach, P. A. (eds.) *A 'Second Edition' of The General Theory*, Vol. 2, pp. 222-243, London: Routledge.

Runde, J. (1998) Clarifying Frank Knight's discussion of the meaning of risk and uncertainty, *Cambridge Journal of Economics*, 22(5), pp. 539-546.

Rutherford, M. C. (1984) Rational expectations and Keynesian uncertainty: a critique, *Journal of Post Keynesian Economics*, 6(3), pp. 377-387.

Samuelson, P. A. (1969) Classical and Neoclassical Theory, in Clower, R. W. (ed.), *Monetary Theory*, London: Penguin.

Savage, L. (1954) *The Foundations of Statistics*, New York: Wiley.

Schumpeter, J. A. (1984) The Meaning of Rationality in the Social Sciences, *Journal of Institutional and Theoretical Economics*, 140(4), pp. 577-593.

Shackle, G. L. S. (1984) Comment on the papers by Randall Bausor & Malcolm Rutherford, *Journal of Post Keynesian Economics*, 6(3), pp. 388-393.

Simon, H. A. (1976) From Substantive to Procedural Rationality, in Spiro J. Latsis (ed.) *Method and Appraisal in Economics*, pp.129-148, Cambrigde: Cambridge University Press.

Solow, R. M. (1985) Economic History and Economics, *American Economic Review*, Papers and Proceedings, 75(2), pp. 328-31.

Tarshis, L. (1980) Post-Keynesian Economics: A Promise that Bounced? *American Economic Review*, Papers and Proceedings, 70(2), pp. 10-28.

Veblen, T. B. (1924[1899]) *The Theory of the Leisure Class*, London: George Allen & Unwin.

von Mises, L. (1928) *Probability, Statistics & Truth*, 2nd revised English edition, Allen & Unwin.

von Mises, L. (1944) The Treatment of "Irrationality" in the Social Sciences, *Philosophy and Phenomenology Research*, 4(4), pp. 527-546.

von Neumann, J. & Morgenstern, O. (1947) *Theory of Games and Economic Behaviour*, Princeton (NJ): Princeton University Press.

Walters, B. & Young, D. (1997) On the Coherence of Post-Keynesian Economics, *Scottish Journal of Political Economy*, 44(3), pp. 329-49.

Watkins, J. W. N. (1952) Ideal Types and Historical Explanation, *British Journal for the Philosophy of Science*, 3(9), pp. 22-43.

Whitehead, A. N. & Russell, B. (1925) *Principia Mathematica*, 2nd edition, Cambridge: Cambridge University Press.

Winslow, E. G. (1986) Human logic and Keynes´s economics, *Eastern Economic Journal*, 12(4), pp. 413-30.

Winslow, E. G. (1989) Organic Interdependence, Uncertainty & Economic Analysis, *Economic Journal*, 99(398), pp. 1173-1182.

Yellen, J. L. (1980) On Keynesian Economics & the Economics of the Post-Keynesians, *American Economic Review*, Papers and Proceedings, 70(2), pp. 15-19).

Notes

[1] These two criticisms do appear, in different forms, in most of the critical surveys of Post-Keyensian (PK) economics which have been published over the last decades. Some examples are Coddington (1976, 1982), Yellen (1980), Tarshis (1980), Walters & Young (1997), and Kakarot-Handtke (2012). In turn, the responses of PK economists to these criticisms are, for instance, in Eichner & Kregel (1975), Hamouda & Harcourt (1988), Crotty (1980, 1994), and Lavoie (2014).

[2] In turn, this was based on Moore´s conception of the real existence of non-natural (or non-physical) entities which Keynes took on board when developing his theory of probability in the *Treatise*. The belief in the real existence of such non-physical entities is known as Platonism. As noted in Bateman (1991, p. 105), `Moore believed in good, and Keynes in probabilities, in exactly the same sense as Plato believed in universal ideal forms´.

[3] This is explained by Keynes as follows:

`… in the sense important to logic, probability is not subjective. It is not, that is to say, subject to human caprice. A proposition is not probable because we think it so. When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion. The Theory of Probability is logical, therefore, because it is concerned with the degree of belief which it is *rational* to entertain in given conditions´ (Keynes, 1952[1920], p. 4).

[4] In a similar vein, Dow (2009, p. 14) points out that Keynes´s *Treatise* addressed the same question that Hume raised with his `problem of induction´.

[5] Carabelli (1988) uses the term `ordinary logic´ to refer to `human logic´ whereas Winslow (1986) uses the latter term which was originally coined by Ramsey (1978[1931]).

[6] According to Lawson (1985), Carvalho (1988), Carabelli (1988) and O´Donnell (1990) this shift did not require a significant change in his basic theoretical framework, however. By contrast, Andrews (1999), Bateman (1987, 1990, 1991) and Davis (1994) hold the opposite thesis.

[7] Similarly, Gillies (2000) has recently propounded an inter-subjective interpretation of probability as an additional version of the subjective epistemological interpretation of probability.

[8] Popper (1990, p. 12) makes it clear that statistical averages will exhibit a tendency to remain stable *only* if the physical conditions remain stable. He makes this phenomenon the basis of his 'propensity theory of probability' (Popper, 1959) which is an *objective* interpretation of the theory of probability. Popper (1990, p. 8) associates the subjectivist theory of probability in the field of physics to Heisenberg and Einstein but he makes clear that he adopts an *objectivist* theory. According to this interpretation, propensities are real forces that stem from physical realities. What is even more important, he recognizes that in our changing real world propensities change *all the time*:

'With the introduction of propensities, the ideology of determinism evaporates. Past situations, whether physical or psychological or mixed, do not determine the future situation. Rather, they determine changing *propensities that influence future situations without determining them in a unique way*… Quite apart from the fact that we do not know the future, the future is *objectively not fixed*. The future is *open: objectively open*' (Popper, 1990, pp. 17-18).

[9] It could be argued that it is the absence of ends-rationality that prevents the notion of logical probability relations from providing a full-fledged theory of human behaviour.

[10] According to O´Donnell (1990), the picture that emerges from Keynes´s analysis in the *General Theory* is also of a 'two-dimensional' and 'two-domain' type of analysis. The two dimensions are expectations and confidence whereas the two domains are *short-term* and *long-term expectations* (O´Donnell, 1990, p. 260). In the first domain, short-term expectations are typified by uncertainty in the context of relatively high weight and confidence whereas, in the second domain, long-term expectations are characterised by uncertainty coupled to either high or low confidence depending upon the situation. The expectations-confidence pair in the *General Theory* is, according to O´Donnell (*op. cit.*), a generalized version of the probability-weight pair in the *Treatise*. This leads him (*op. cit.*, p. 263) to contend that Keynes advanced a pioneering *theory of rationality under radical uncertainty* consisting of two independent dimensions with economic agents forming expectations about outcomes and degrees of confidence about the latter.

[11] O´Donnell (1990, p. 259) notes that the `continuity vs. change´ issue in Keynes´ philosophical views after 1921 has led to a division of opinion. On the one hand, there are those like Winslow (1986, 1989) or Davis (1994) who argue that in 1932, and under the strong impact of Ramsey´s (1978[1931]) criticism of Keynes´s later views, Keynes switched to a non-logical theory of probability. In a similar vein, Shackle (1984, p. 391) writes that 'Keynes the economist abandoned probability in any technical form… Keynes´s probability theorizing has no overt bearing on the economics which he published'. On the other hand, there are those like Lawson (1985), Rutherford (1984, p. 381), or Carvalho (1988, p. 72) who argue that Keynes did not adopt an alternative conception of probability but rather continued to work within the theoretical framework of the *Treatise on Probability*. Other authors adopt a middle-ground position. For instance, O´Donnell (*op. cit.*) argues that Keynes´s thought continued to be grounded on the conceptual framework

of the *Treatise* albeit there was an internal change of emphasis within his framework after 1932 which consisted of an increase in the relative significance of the 'indeterminate' domain and 'weak' rationality. Likewise, Runde (1997, p. 241) maintains that, in his later work, Keynes gave up his ontology of logical probability relations albeit he continued to regard qualitative probability comparisons as basic to epistemic probability.

[12] In this respect, Andrews (1999, p. 10) argues that these two notions have outlasted other elements of the *Treatise* because they are not crucial to its central argument. As we show below, its central argument was rejected by Keynes in the wake of Ramsey´s critique of it.

[13] According to Runde (1990), the dichotomy between probability and confidence (and weight) can also be extended over to Keynes´s distinction between `risk´ and `liquidity´ premium alluded to in a letter addressed to Townshend and written after the publication of the *General Theory*:

'I am rather inclined to associate risk premium with probability strictly speaking, and liquidity premium with what in my *Treatise on Probability* I called "weight". An essential distinction is that a risk premium is expected to be rewarded on the average by an increased return at the end of the period. A liquidity premium, on the other hand, is not even expected to be so rewarded. It is a payment, not for the expectation of increased tangible income at the end of the period, but for an increased sense of comfort and confidence during the period' (Keynes, 1973b, reproduced in Runde, 1990).

Be that as it may, a reading of Keynes´ writings suggests that he does not view `confidence´ and `weight´ as being interchangeable albeit he would probably concede that they tend to move in the same direction in which case the appropriate conception of weight is 'some [subjective] measure of the degree of completeness of the information on which a decision is based' (Runde, 1990, p. 287). Thus, if we take expression (2) above as the most adequate interpretation of the notion of weight, then the greater the weight of evidence in favour of a certain forecast the more robust the basis on which to formulate it and, hence, the more confident we will be that our forecast is an appropriate guide to action.

[14] The notion of `weight of the argument´ also appears in the *General Theory* in the context of Keynes´ discussion of investment decision-making.

[15] However, Rutherford (1984, p. 379-80) argues that 'Keynes contended that even a large quantity of favorable evidence could not logically impart a high probability to a hypothesis if the evidence emanated from repetitions of *identical* experiments or observations made under *identical* conditions' (Keynes, 1952[1920], pp. 217-219). According to him, it follows from this that a certain amount of evidence will mean more (less) if it is collected under varied (similar) conditions.

[16] For instance, Dow (1995, p. 726) points out that, according to Keynes, `more weight [should] be given to facts of which we are more certain than to facts of which we are very uncertain even if they are more relevant to the argument´.

[17] Similarly, Dow & Dow (2011, pp. 6-7) argue that `the degree to which ignorance is recognised is ultimately a matter of psychology´ and, hence, they implicitly suggest that the `weight of an argument´ is subjective.

[18] In particular, and referring to the notion of `weight´, Keynes writes that `the question to be raised in this chapter is somewhat novel; after much consideration I remain uncertain as to how much importance to attach to it´ (Keynes, 1952[1920], p. 71).

[19] Keynes alludes to `weight´ in the *General Theory* (1936, p. 148) and in a letter to Townshend written in 1938 (Keynes, 1973*b*, p. 293) where Keynes relates `weight´ to the notion of `liquidity premium´. In the *General Theory*, Keynes gives it an explanatory role in three different areas: uncertainty, confidence, and liquidity preference. Specifically, `weight´ exerts two effects on the current level of investment. First, it influences the marginal efficiency of capital through its impact on the level of confidence. Second, it influences the rate of interest through its relation to liquidity preference.

[20] Savage´s theory builds upon the work of both Ramsey (1978[1931]) and De Finetti (1937) who showed that subjective probability could be defined in terms of preferences over gambles and of von Neumann and Morgenstern (1947) who provided the first axiomatic derivation of expected utility theory. Another

example is Ramsey´s optimal growth model (Ramsey, 1978[1931], ch. 11) which represents the standard approach to analysing optimal growth in macroeconomics.

[21] Likewise, in his critique of Keynes´s notion of *a priori* probability relations, Ramsey (1978[1931]), p. 94) suggests that *a priori* probabilities would be `determined by natural selection´. However, it has been suggested that one of the reasons Keynes failed to provide an alternative to the frequency approach to probability in the *Treatise* was that `he made use of an epistemology of intuition that lacked a theory of error´ (Davis, 1994, p. 102). We argue below that Keynes adopted in his later economic writings a theory of human behaviour based on conventional judgement which can, in turn, be provided with a trial and error-elimination mechanism in a way that makes the former compatible with evolutionary epistemology.

[22] However, Ramsey´s notion of `human logic´ implies, according to Keynes, more than mere consistency of thought:

`[Ramsey] was led to consider "human logic" as distinguished from "formal logic". Formal logic is concerned with nothing but the rules of consistent thought. But in addition to this we have certain "useful mental habits" for handling the material with which we are supplied by our perceptions and by our memory and perhaps in other ways, and so arriving at or towards truth; and the analysis of such habits is also a sort of logic´ (Keynes, 1973b, p. 338)

[23] See, for instance, Winslow (1989, p. 1178).

[24] Keynes implicitly accepted the validity of the frequency approach to probability in the natural sciences in a letter addressed to Harrod on 4th July 1938 the full content of which is shown in a footnote below.

[25] This may become clear after reading a fragment of the discussion where he introduces his framework:

`Let us give an instance of the sort of case which might occur. I am at a cross-roads and do not know the way; but I rather think one of the two ways is right. I propose therefore to go that way but keep my eyes open for someone to ask; if now I see someone half a mile away over the fields, whether I turn aside to ask him will depend on the relative inconvenience of going out of my way to cross the fields or of continuing on the wrong road if it is the wrong road. But it will also depend on *how confident I am that I am right*; and clearly the more confident I am of this the less distance I should be willing to go from the road to check my opinion. I propose therefore to use the distance I would be prepared to go to ask, as a measure of the confidence of my opinion...´ (Ramsey, 1978 [1931], pp. 76-77, emphasis added).

Indeed, Ramsey´s notion of `degree of belief´ is precisely the *measure* of the confidence of one´s opinion and is, thus, subjective. In any case, the paragraph highlights that the distance I am willing to go out of my way to check my opinion is also the maximum distance I *should be willing* to go out of my way to check my opinion if I really wished to realize the objects of my desire. We believe the example shows that *Ramsey´s framework is an example of RPs* in that `the theoretician reconstructs P-S as she believes the agent believes it is´.

[26] This is nicely explained in Davis (1994):

`In Keynes´s later career, then the effort to explain the operation of the economic world produced a change in philosophical thinking that forced abandonment, modification, and replacement of much that Keynes had previously believed. Intuition in the Moorean sense was replaced by individual expectation. The focus on probability became secondary to the focus on convention. Rational behaviour as a principal concern in the analysis of individual judgement was supplanted by a preoccupation with the effects of interdependence and uncertainty… Unfortunately, Keynes never clearly articulated his philosophical conversion, largely no doubt on account of the tremendous demands upon his time made by economics and policy´ (Davis, 1994, pp. 146-7).

[27] Choi (1993, pp. 63-66) provides some insightful ideas stemming from psychology which enhance the Keynesian argument that economic agents rely on conventional `rules of thumb´ when making decisions. For instance, he argues that our confidence in the way we do things tends to be reinforced by the approval of others as well as by our belief that our own behaviour merits the applause of others.

[28] Runde (1997, p. 228) argues that should Keynes have written a 'second edition' of the *General Theory* he would have elaborated on the theme of conventional methods of expectations formation as surrogates for mathematical calculus and the way in which such methods contribute to financial market instability.

[29] Keynes´s insight is eloquently put in Iwai (2009, p. 8, emphasis added):

`In the end, the only reason a particular face is selected as the prettiest is that every competitor [in the Beauty Contest] believes every other competitor believes she is selected as the prettiest, without any support from reality, either objective or subjective. *The prettiest is the prettiest merely because she is selected as the prettiest*. What we see here is the working of the "bootstrapping" logic of Baron Münchausen who claimed he had pulled himself out of a swamp by pulling on his own bootstraps´.

[30] For instance, Littleboy (1990, p. 29) argues that `one of Keynes´s most important innovations lay in the realization of the significance of conventions that arise when transactors, confronted by an uncertain environment, are psychologically disposed to act in a manner in which they study and imitate the actions of others´. He advances a theory of macroeconomic dynamics grounded on the interaction of different conventions in different spheres of the economic system (*op. cit.*, pp. 289ff).

[31] Davis (1997) argues that the theory of the formation of social conventions in Keynes´ theory remains undeveloped. For instance, he notes the *interactive* character of the formation of conventional judgements and suggests that in a hypothetical 'second edition' of the *General Theory* Keynes would have focused on the interaction between individual and average opinion and on how confidence varies according to the degree of coincidence between them. He discusses several hypothetical scenarios where, depending on the degree of coincidence between individual and *average* opinion, the 'shared' degree of confidence may either vary or remain stable (*op. cit.*, pp. 216-17). However, he does not see an evolutionary dimension in Keynes´s approach to the formation of conventions albeit he writes that 'for Keynes, then, the conditions associated with states of confidence concern the success or lack of success with which individuals come to assess each other´s opinions about markets' (*op. cit.*, p. 217). Notwithstanding it, Howitt (1997, p. 241) points out that it would have been natural for Keynesian economics to follow the approach suggested by behavioural or evolutionary economics in the study of the causal mechanisms underlying the formation of conventional expectations.

[32] Similarly, as noted in Lawson (1985, p. 918), the notion of rationality that stems from Keynes´s macro-theory is compatible with Simon´s notion of 'bounded' rationality.

[33] Arguably, it may not be possible for some individuals to observe *directly* how successful other people have been in the past and, further, the success or lack of it of the latter may not be a direct outcome of the adoption by some people of particular beliefs. As to the former issue, some people may assume that those who have a higher social status have adopted beliefs which tend to be correct on average as in Veblen´s theory of social emulation (Veblen, 1924[1899]). As to the second issue, we need to assume that success is *mainly* associated to the holding of `correct´ beliefs.

[34] In this account of Keynes´s macro-theory there is, arguably, very little scope for `irrational´ behaviour. In some interpretations of the former, `irrationality´ is usually associated to the notion of `animal spirits´. However, as Dow & Dow (2011, p. 7) note, Keynes recognised that `animal spirits´ is less relevant to decision-making than conventional judgement, at least in the context of financial markets. Keynes uses the term `animal spirits´ three times in chapter 12 of the *General Theory* where he refers to the decision by entrepreneurs to invest in real capital:

`Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as a result of *animal spirits* — of a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities… Thus if the *animal spirits* are dimmed and the spontaneous optimism falters, leaving us to depend on nothing but a mathematical expectation, enterprise will fade and die… But individual initiative will only be adequate when reasonable calculation is supplemented and supported by *animal spirits*, so that the thought of ultimate loss which often overtakes pioneers, as experience undoubtedly tells us and them, is put aside as a healthy man puts aside the expectation of death´ (Keynes, 1936, pp. 161-62, emphasis added).

We thus believe that `animal spirits´ plays a marginal role in Keynes´s theory. To us, the role of `animal spirits´ in Keynes´s theory is similar to Hume´s idea that the ultimate driver of human action is not reason but sentiment or passion or as he puts it: 'I shall endeavour to prove first, that reason alone can never be a motive to any action of the will; and secondly, that it can never oppose passion in the direction of the will' (Hume, 1978[1739-40], p. 413). In terms of the conceptual framework used in the previous chapter, `animal spirits´ may only amount to (unconscious) aims of individuals and, as such, it has to be taken as given for the purpose of macroeconomic analysis. In the context of Popper´s *SA*, `animal spirits´ represent an exogenous element in the `problem-situation´ individuals find themselves. Be that as it may, we believe there is very little explanatory and predictive power we can gain by focusing on the role of `animal spirits´ and, hence, that a more promising avenue for research in macroeconomics is to study the conventional basis of economic behaviour.

[35] In this respect, Runde (1997) argues that, in chapter 12 of the *General Theory*, Keynes elaborates on the theme of conventional methods of calculation as surrogates for mathematical calculation and the ways through which conventional behaviour may lead to financial instability. He then conjectures that Keynes would have reinforced this part of his argument had there been a `second edition´ of *The General Theory*.

[36] However, some HC may withstand the passing of time if people keep on believing that they represent useful guides for behaviour. The three conventions Keynes (1937, pp. 214-15) refers to and which were listed above are examples of HC which have successfully withstood the passing of time.

[37] Incidentally, when referring to the notion of 'sunspot' equilibria in neoclassical theory, Arrow (1986) points out that the former refers to the existence of a continuum of equilibria in which one equilibrium is based on fundamentals and the remaining equilibria depend on the contingency that becomes relevant simply because everyone *believes* it is relevant. He adds that, in such cases 'we can have situations where social truth is essentially a matter of *convention*, not of underlying realities' (Arrow, 1986, p. S396, emphasis added).

[38] In particular, and in addition to the issue we address in this section and which concerns the alleged non-homogeneity of economic material over time, Keynes also raises in those letters several issues related to the problem of omitted variables, lag lengths, and model specification in his correspondence with Harrod and Tinbergen.

[39] As Davis (1994, p. 144) writes, `that the future is not a function of "past statistics" and that one must always leave some room for "expectation and the state of confidence relating to the future" suggests that Keynes did not believe that there was much scope for inductive methods in economics´. Likewise, Nobel Laureate in Economics John Hicks declares that 'I am bold enough to conclude from these considerations that the usefulness of "statistical" or "stochastic" methods in economics is a good deal less than is now conventionally supposed' (Hicks, 1979, p. 129).

[40] We reproduce the letters below. The first letter addressed to Harrod was issued on 4th July 1938:

`It seems to me that economics is a branch of logic, a way of thinking; and that you do not repel sufficiently firmly attempts à la Schultz to turn it into a pseudo-natural science. One can make quite worthwhile progress merely by using your axioms and maxims. But one cannot get very far except by devising new and improved models. This requires, as you say, "a vigilant observation of the actual working of our system". *Progress* in economics consists almost entirely in a progressive improvement in the choice of models… But it is of the essence of a model that one does *not* fill in real values for the variable functions. To do so would make it useless as a model. For as soon as this is done, the model loses its generality and its value as a mode of thought… The object of statistical study is not so much to fill in missing variables with a view to prediction, as to test the relevance and validity of the model… Economics is a science of thinking in terms of models joined to the art of choosing models which are relevant to the contemporary world. It is compelled to be this, because, unlike the typical natural science, the material to which it is applied is, in too many respects, not homogeneous through time. The object of a model is to segregate the semi-permanent or relatively constant factors from those which are transitorily or fluctuating so as to develop a logical way of thinking about the latter, and of understanding the time sequences to which they give rise in particular cases… Good economists are scarce because the gift for using "vigilant observation"

to choose good models, although it does not require a highly specialised intellectual technique, appears to be a very rare one´ (Keynes, 1973b, pp. 295-7)

The second letter addressed to Harrod was issued on 16th July 1938:

`The point needs emphasising because the art of thinking in terms of models is a difficult — largely because it is an unaccustomed — practice. The pseudo-analogy with the physical sciences leads directly counter to the habit of mind which is most important for an economist proper to acquire. I also want to emphasise strongly the point about economics being a moral science. I mentioned before that it deals with *introspection and with values. I might have added that it deals with motives, expectations, psychological uncertainties*. One has to be constantly on guard against treating the material as constant and homogeneous. It is as though the fall of the apple to the ground depended on the apple´s motives, on whether it is worth while falling to the ground, and whether the ground wanted the apple to fall, and on mistaken calculations on the part of the apple as to how far it was from the centre of the earth´ (Keynes, 1973b, pp. 299-300, emphasis added).

[41] Although Popper apparently borrowed the terms `explanation in principle´ and `explanation in detail´ from Hayek (1967), to the best of our knowledge, these terms are mentioned for the first time in Watkins (1952).

[42] A classic discussion of this topic in the context of economics is in Papandreou (1959). He argues that (i) economists construct models rather than theories, and (ii) economic models are *strictly explanatory devices*. More precisely, he argues that economic models cannot be used as predictive devices because the conditions of their applicability cannot be set out in advance (*op. cit*., p. 1099).

[43] In the literature on Keynesian economics, the non-homogeneity of economic data is usually associated to the mathematical notion of non-ergodicity. According to Davidson (1996, p. 479), most mainstream economists implicitly assume the existence of a predetermined, immutable or ergodic world that can be fully described by objective conditional probability distributions. He identifies three different decision-making economic environments: (i) the objective probability environment, (ii) the subjective probability environment, and (iii) the *truly* uncertain environment (Davidson, 1991). In the objective probability environment, the realm of the `rational expectations hypothesis´ (REH), decision-makers believe that the past is a statistically *unbiased* guide to the future. As Davidson explains, for the REH to provide a theory of expectations formation, not only must the subjective and objective probability distribution functions coincide at any point in time but they must also be derived from 'ergodic' stochastic processes. The latter exhibit the key property that 'averages calculated from past observations cannot be persistently different from the time average of future outcomes' (*op. cit*., p. 132). Non-ergodicity is sometimes likened to non-stationarity. However, as Davidson (*op. cit*) makes clear, non-stationarity is a sufficient, but not necessary condition, for non-ergodicity.

[44] More specifically, Robbins argues that the ultimate premises of social science are human dispositions. As he explains, such dispositions `are so much the stuff of our everyday experience that they have only to be stated to be recognised as obvious´ (Robbins, 1932, p. 79).

[45] Notwithstanding it, this is not necessarily the case in practice. For instance, Heiner (1983, p. 561) notes that (neoclassical) optimization models are unable to imply the `Law of Demand´ (i.e., that an increase in the relative price of a commodity will lead to a decrease in its demand) which is, arguably, the simplest empirical regularity in economics. He notes that we can certainly use neoclassical consumer theory to argue that it is *unlikely* that a negative income effect will outweigh the substitution effect yet we cannot be sure that this will be the case. Notably, he comments the following:

`I was told in a graduate price-theory class by Armen Alchian that the only clear implication of consumer theory is that with more income, a consumer will buy more of at least something. Harold Demsetz, when informed of this story, responded by saying, "well then just define holding cash balances as saving, and we have no testable implications, just one mass of tautologies"´ (*op. cit*., footnote 4).

[46] Of course, this presupposes that the statistical environment is stable. As we have explained above, the mathematical expression of this idea is represented by the notion of ergodicity. Incidentally, the latter is

identified by Nobel Laureate Paul Samuelson (1969, p. 184) as the 'sine qua non of the scientific method in economics'.

[47] Both Keynes (1920[1952]) and Knight (1971[1921]) associate the notion of uncertainty with non-measureable probability. We have already referred above to the definition of uncertainty in Keynes´s work. As for the definition of uncertainty in Knight (*op. cit.*), Runde (1998, p. 543) argues that Knight proposes a *tripartite* schema of probability which can be restated as follows:

1. 'Classical' or *a priori* probability which corresponds to the ideal case in which numerical probabilities can be computed by assigning them to equally probable and mutually exclusive possible outcomes such as the six sides of a perfect die.
2. 'Statistical' probability which refers to situations in which frequencies may be obtained by carrying out an empirical (and complete) taxonomy of potential outcomes which are then divided into classes of *less* than perfectly homogenous trials.
3. 'Estimates': this corresponds to situations in which it is either impossible to calculate *a priori* probabilities or where there is an *insufficient* number of trials which are 'like' enough to construct a reference class of trials in order to obtain frequencies.

As Runde (*op. cit.*) notes, this reformulated version of Knight's taxonomy of probability situations leads to the key Knightian distinction between `risk´ and `uncertainty´. The former refers to situations in which decision-makers can calculate either *a priori* or statistical probabilities whereas the latter refers to situations in which it is not possible to obtain either *a priori* or statistical probabilities. Crucially, Knight makes it clear that, in practice, it is impossible to obtain an entirely homogenous taxonomy of different types of trials thereby making the theoretical distinction between 'statistical' probability and 'estimates' a *matter of degree* (*op. cit*., p. 225).

[48] We argued above that one of the themes of the *Treatise on Probability* that made its way into Keynes´s later economic writings is the possibility that probability is *unmeasurable*. Therefore, most interpretations of the notion of `uncertainty´ in Keynes tend to associate uncertainty with situations in which we cannot measure the relevant probabilities. Keynes refers to this scenario in his *QJE* paper (Keynes, 1937). For instance, he argues that entrepreneurs cannot rely on any 'scientific' evidence when making long-term investment decisions because most investment decisions are *unique* in the sense that any statistical data generated from the past is not useful to evaluate the statistical probability of different potential outcomes. As he explains in a much-quoted passage:

'By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty; nor is the prospect of a Victory bond being drawn. Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth-owners in the social system in 1970. About these matters there is *no scientific basis* on which to form any calculable probability whatever. We simply do not know' (*op. cit*., pp. 213-14, emphasis added).

[49] By contrast, in the theoretical framework developed in Heiner (1983, p. 561), the predictable features of individual behaviour do not arise from optimizing with no uncertainty in choosing their most preferred behaviour. On the contrary, Heiner shows that, in the special case of no uncertainty, the behaviour of fully optimizing agents who respond with complete flexibility to every perturbation in their environment would not produce easily recognizable patterns but rather would be extremely difficult to predict.

[50] An example of this is the story in John Ford's classic Western `The Man Who Shot Liberty Valance´. Ranse Stoddard (James Stewart's character) makes a career as a State Governor and U.S. Senator largely on the basis of his reputation as `the man who shot Liberty Valance´. However, as we all know, it was Tom Doniphon (John Wayne´s character) who really shot him.

[51] As Hempel (1949, p. 467, emphasis added) argues:

`The method of empathy is, no doubt, frequently applied by laymen and by experts in history. But it does not by itself constitute an explanation; it is rather essentially a *heuristic device*; its function is to

suggest certain psychological hypotheses which might serve as explanatory devices in the case under consideration´.

[52] No doubt, the ability of Keynes to adopt the point of view of financial market participants in this case was, arguably, enormously facilitating by his own experience as a speculator in London´s stock market.

[53] In the *General Theory* Keynes (1936, p. 96) defines the `psychological law of consumption´ as the law according to which `men are disposed, as a rule and on average, to increase their consumption as their income increases, but not by as much as the increase in their income´.

[54] This account of the conduct of investors is, according to Runde (1990), consistent with the terms of the discussion in Chapter 26 of the *Treatise* where Keynes distinguishes between 'the most probable forecast we can make and our confidence in that forecast' (Keynes, 1952[1920], p. 286) and where he links low `weight´ to low confidence.

[55] That a decrease in `confidence´ and its associated increase in liquidity preference constitute a `rational´ response of economic agents is hardly questionable. For instance, in many choice situations more can be learned about the factors governing eventual outcomes after decisions have been made. This presents no problem where choices can be completely and costlessly reversed. But things are very different where they cannot. Such irreversibility entails that it may sometimes be `rational´ to suspend judgement and delay commitment until more information has been acquired. In this context, the function of liquidity is that of `giving us time to think´ (Hicks 1974, p. 57). In general, however, decisions cannot be postponed until all the evidence is available and, in such cases, the readiness and terms at which commitments to previous choices may be dissolved assumes importance (Runde 1994, p. 136). Therefore, in situations of uncertainty liquid assets carry a premium over illiquid ones.

[56] For instance, Carvalho (1988, p. 79) distinguishes between long-term investment decisions and short-term production decisions. Unlike the former, short-term production decisions tend to be repetitive, do not normally imply irreversible commitments of financial resources and can be checked and reversed after very short intervals.

III. CONTEMPORARY MACROECONOMICS AND POPPER´S THEORY OF KNOWLEDGE AND LEARNING

`No period of history has ever been great or ever can be that does not act on some sort of high, idealistic motives, and idealism in our time has been shoved aside, and we are paying the penalty for it.´ (A. N. Whitehead, in *Dialogues of Alfred North Whitehead,* ch. 32, 1954)

1. Introduction

According to Woodford (2009) there was a substantial convergence of views in macroeconomics over the last two decades. The evidence for such convergence is the observation of a cessation of methodological struggles within macroeconomics. More specifically, he recognizes that a New Synthesis emerged in macroeconomics over the last two decades or so. A key component of the New Synthesis is the adoption by most macroeconomists of the 'Rational Expectations Hypothesis' (REH) as their usual way of modelling economic agents´ expectations formation. Likewise, Goodfriend & King (1997) argue that a 'New Neoclassical Synthesis' model emerged in macroeconomics in the 90s and note that REH is a key ingredient of it. More recently, Farmer (2013) has recognized that REH is a useful theoretical device which, if applied carefully, can help us understand what went wrong in the last financial crisis. This stands in stark contrast to those theorists who believe that contemporary macroeconomic and financial theory and, particularly REH, bears some responsibility for the onset of the last financial crisis (Guesnerie, 2013, p. 50). Recent critical assessments of REH add to criticisms made in the past as expounded by a number of scholars including several Nobel Prize Awards in

economics like, for instance, Arrow (1978), Tobin (1981), Krugman (2009), and Stiglitz (2011). Criticisms of REH tend to focus on the notion that it begs epistemological issues such as how market participants acquire the knowledge they need to produce 'rational expectations' (RE). For example, Frydman & Goldberg (2013, p. 130) argue that 'REH has no connection to how even minimally reasonable profit-seeking individuals forecast the future in real world markets'. In a series of contributions, Frydman & Goldberg (2007, 2011, 2013) argue that economists´ belief in the efficacy of REH stems from the premise that they have discovered a universally valid way to capture how individuals make decisions. However, this presumption has profound methodological implications. First, fallibility is restricted to economic agents´ inability to forecast exogenous shocks. Second, models which fully pre-specify how the surrounding environment unfolds over time obliterate agents´ creativity. As noted in Phelps (2007, p. xv), it is *contradictory* to adopt the REH premise that whatever change takes place in the future is either known or knowable in the present since —as Schumpeter argued— change and innovation are key features of capitalist economies[1]. Third, by imposing onto macroeconomic models the restriction that economic agents´ expectations coincide on average with the conditional forecast generated by the theorist´ model it turns out that fully specified macroeconomic models rule out the possibility that agents´ beliefs *independently* affect outcomes in a way that model builders cannot fully pre-specify. Another way of saying this is that REH implies that, to the extent that agents´ beliefs are the product of macroeconomic theory, forecasting strategies are determined *jointly* at the individual and aggregate levels. As a result of it, it has been argued that macroeconomic models that adopt REH lack plausible micro-foundations (Frydman & Goldberg, 2013). Lastly, it may well be 'irrational' to behave as predicted by REH if market participants do not know the 'true' model and

know, in turn, that nobody knows it. Thus, the question emerges as to what the appropriate notion of rationality is in macroeconomic models if REH is abandoned.

Now, we believe that the contradictions embedded in REH and the NNS model stem mainly from their flawed epistemological and ontological foundations. Therefore, if a substantial reorientation of macroeconomics is to take place in the future, the first task to address is to provide it with solid epistemological and ontological foundations. In view of this, the main claim we want to make is that Popper´s evolutionary theory of knowledge and learning (PTKL) offers such solid foundations. In particular, and unlike macroeconomic models grounded on REH, PTKL presupposes that the 'true' model of the economy is *unknown* by economic agents and theorists alike and presumes instead that: (i) all knowledge is conjectural, (ii) the adaptation of knowledge to the surrounding environment is always and everywhere imperfect, and, hence, never optimal, and (iii) economic agents´ decisions and actions may well bring about irreversible changes in the surrounding environment which they may not be able to fully anticipate. Further, PTKL implies that the main source of change in a complex adaptive system such as a market economy is the interaction between economic agents´ conjectures and their subsequent revisions in the face of observed discrepancies between expected and realized outcomes. In short, we argue below that PTKL may provide contemporary macroeconomics with robust epistemological and ontological foundations by allowing the former to take full account of human fallibility.

The content of this essay is as follows. The following section contains a review of REH. We analyse REH in the light of PTKL and identify their differences in section 3. Since PTKL was reviewed in chapter 1 above, we refer readers to that chapter if need be. We explore the main elements of a new conceptual framework for macroeconomics grounded upon PTKL in section 4. The focus of our discussion will be on endogenous

change, the ephemeral nature of macroeconomic equilibrium, the `partial´ autonomy of expectations, and the notion of rationality. Finally, section 5 summarizes and concludes.

2. The Rational Expectations Hypothesis

No wonder REH has been the dominant approach to modelling economic agents´ expectations in macroeconomics over the last decades. REH has been interpreted as an attempt to provide macroeconomics with a theory of expectations formation that is *a priori* consistent with the optimization hypothesis. We owe its first formulation to Muth (1961) who suggests that expectations should be modelled in a way that allows them to change *endogenously* when the structure of the system changes:

'I should like to suggest that expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory… The hypothesis can be rephrased a little more precisely as follows: that expectations of businesses (or, more generally, the subjective probability distribution of outcomes) tend to be distributed, for the same information set, about the prediction of the theory (or the "objective" probability distributions of outcomes)' (Muth, 1961, p. 316).

Therefore, according to Muth (1961, p. 316), REH implies that economic agents´ subjective expectations are, on average, equal to the `true´ values of the variables. In other words, it is only the *average* of economic agents´ forecasts that will be equal to the mathematical expectation of the variable. As a result of it, the forecast of a *given* individual may not coincide with the latter.[2] Yet, as Knudsen (1993, p. 153) observes, the reintroduction of the REH by Lucas entailed a reinterpretation of it. In particular, REH

was now reinterpreted as implying that the expectations of *every* single agent were correct on average. The reason Lucas (2001, p. 13) characterizes REH this way is that, under the premise that economists know the 'true' model of the economy, he interprets the systematic forecast errors made by those market participants who do not formulate their expectations according to REH as a symptom of their irrationality. Accordingly, the presence of significant but nevertheless unexploited correlations between market participants´ forecast errors and the information available to them was now interpreted as pointing to unrealized profit opportunities in real-world markets. As Lucas (2001, p. 13) puts it, 'if your theory reveals profit opportunities, you have the wrong theory'. This means that, in addition to the premise that the 'true' model of the economy is known or knowable, REH imposes two further requirements on economic models: 'substantive' rationality by *all* market participants and mutual consistency of market participants´ expectations.[3] For the purpose of this essay, a key issue is that REH is viewed by one of its most qualified proponents as requiring a high degree of stability and regularity over time of the phenomenon under study:

'Evidently, this hypothesis [Muthian rationality] will not be of value in understanding psychotic behaviour. Neither will it be applicable in situations in which we cannot guess which, if any, observable frequencies are relevant: situations which Knight called "uncertainty". It will most likely be useful in situations in which the probabilities of interest concern a fairly well defined *recurrent* event, situations of "risk" in Knight´s terminology… In cases of uncertainty, economic reasoning will be of no value… ' (Lucas, 1977, p. 15, term in brackets and emphasis added).

Thus, REH presumes that phenomena exhibit sufficient stability and regularity over time to allow economic agents to infer their stylized facts and attach probabilities to an exhaustive list of potential outcomes. In this respect, Lucas' position coincides with that of Savage (1954), the father of subjective expected utility (SEU) theory, who defines 'small worlds' as situations of perfect knowledge where all alternatives, their consequences, and their probabilities are known for certain. Examples of 'small worlds' are hazard games such as lotteries and roulette. According to Savage (*op. cit*.), these are the only types of environments in which the application of Bayesian decision theory is legitimate. By contrast, he defines 'large worlds' as those environments where part of the needed information is missing or else where the future is 'uncertain' in the sense of Keynes (1920) and Knight (1971[1921]). Crucially, Savage (1954, p. 16) insists that applying Bayesian theory to decisions in 'large worlds' does not make any sense because there is simply no way to know all the alternatives, their consequences, and their probabilities. Therefore, the regularity and predictability of economic phenomena that Lucas (*op. cit*.) presents as the *sine qua non* condition for the emergence of (presumably neoclassical) economic reasoning requires that the environment is stochastically stable or `ergodic´ so that market participants can extrapolate into the future the empirical regularities inferred from the past. In particular, the latter exhibit the key property that averages calculated from past observations can persistently differ from the time average of future outcomes (Davidson, 1991, p. 132).[4] As Davidson (*op. cit*.) explains, for the REH to provide a theory of expectations formation, not only must the subjective and objective probability distribution functions coincide `at any point in time for all market participants´ but they must also be derived from ergodic stochastic processes. Thus, Lucas (1977) apparently accepts that some economic phenomena are non-ergodic but he suggests that economics should be concerned *only* with ergodic worlds.

Next, Lucas´ allusion in the quotation above to 'repeated instances of essentially similar events' also suggests that, at least initially, he had in mind an inductive theory of knowledge acquisition or learning. However, an inductive learning theory is necessarily grounded on a 'cumulative' theory of knowledge. The latter implies that, as the amount of information collected increases over time, the probability that our hypothesis is true increases and converges asymptotically to unity. Yet, as Boland (2003) argues, both the 'cumulative' theory of knowledge and the inductive theory of learning are incompatible with Popper's philosophy. To be sure, according to Popper all knowledge is conjectural and, hence, no matter how much (finite) favourable empirical evidence we accumulate, the likelihood of a hypothesis being `true´ does not increase. In any case, Lucas (1986) apparently adopted a somewhat different theory of learning later on. Specifically, when seeking to justify the optimization assumption he characterizes the situations on which economic theory focuses as the end-result of an *adaptive* learning process:

'Economics has tended to focus on situations in which the agent can be expected to "know" or to have learned the consequences of different actions so that his observed choices reveal stable features of his underlying preferences… Technically, I think of economics as studying decision rules that are steady states of some adaptive process, decision rules that are found to work over a range of situations and hence are no longer revised appreciably as more experience accumulates…' (Lucas, 1986, p. 218).

This quotation reveals some features of Lucas's later methodological approach which are not present in Lucas (1977). First, he seems to adopt a 'cumulative' theory of knowledge as he refers to the accumulation of experience. Second, he suggests that the decision rules neoclassical economic theory deals with, namely, those decisions rules that

stem from the implementation of optimization methods can be interpreted as steady states of an *adaptive* learning process. Although he does not elaborate on it, the latter presumably consists of a trial and error-elimination process whereby economic agents make decisions which may turn out to be right or wrong so that by discarding wrong decision rules and retaining the right ones they eventually 'learn'. Such learning process bears some resemblance to PTKL in the sense of apparently being *adaptive*. However, as we argue below, such resemblance is superficial. Third, Lucas (1977) assumes that the learning process eventually converges to an optimum. Specifically, he observes that those forecasting rules which are not based on REH will tend to generate systematic forecast errors thereby revealing the existence of *unrealized* profit opportunities, the implicit proposition being that if forecasting rules are based upon REH an optimum will be attained and unrealized profit opportunities will thus not emerge (Lucas, 2001). Yet, as we argue below, this represents an extreme position which requires that, at least, the two following conditions be fulfilled: (i) the object of knowledge (i.e., an economy) must exhibit stability overtime and (ii) the feedback mechanism that connects economic agents´ expectations and realized outcomes must be fast and accurate enough. Finally, and crucially, Lucas appears to suggest that economics should be concerned solely with those decision rules that economic agents have adopted *after* the 'true' model has been found, the implicit assumption being that decisions made *prior* to its discovery do not have any impact on the economy, i.e., they do not affect the equilibrium. However, this logically requires either that the 'true' economic model exists *prior* to decisions made by economic agents or else that the latter´s actions do not affect the economy. Again, this feature of the 'learning' process alluded to in Lucas (1986) is closely linked to the ergodicity assumption referred to above in that a pre-existing reality that is independent of agents´ beliefs and actions is only compatible with an ergodic environment.

3. The `Rational Expectations Hypothesis´ in the light of Popper

In this section we examine REH in the light of PTKL. Our purpose is to identify the key differences between them in order to substantiate the claim that PTKL provides more solid epistemological and ontological foundations upon which macroeconomists can build a more useful macro-theory. In particular, we argue below that REH diverges dramatically from PTKL in, at least, three respects: the learning process, the nature of knowledge, and the nature of the world and human action.

3.1. Can we ever get to know the `true´ model?

The main contradiction REH encounters when viewed in light of PTKL is that the former does not possess a mechanism that allow economic agents to get to know the 'true' model. As we saw in chapter 1, PTKL implies that we can *never* know whether a certain hypothesis is 'true' and, hence, the assumption that economic agents know the 'true' model is fallacious. First, if we follow Popper, no matter how many models are tested and subsequently replaced, there is no way we can know whether the model we use for predictive purposes is 'true'. By the 'true' model we mean, of course, the model that produces unbiased estimates. This is not only because there is not an `inductive logic´ but also because, as Popper explains, any model is always an *over-simplification* and, hence, false (Popper, 1994, p. 166). Consequently, all we can obtain is a model that produces 'better' estimates than other models in a *particular* historical and institutional context — and even this may be hard to achieve due to the practical problems posed by the Duhem-Quine thesis — but this does not prove that our model is 'true'.[5] In other words, we can never prove that a model is 'true'. Thus, REH is a misnomer. It is not a hypothesis; it is an assumption.

Second, let us assume for the sake of convenience that economic agents happen to know the 'true' model of the economy so that their forecasts are unbiased on average and their errors are random, i.e., they fall on a bell-shaped curve around the 'true' value. Now, if all errors are assumed to be random, then economic agents will not be able to 'learn' from them in the sense that they will not be able to revise their knowledge in the aftermath of a change in the surrounding environment that brings about a corresponding change in the 'true' model. To be sure, *PTKL implies that we 'learn' by eliminating our non-random errors*. In other words, error-elimination requires that realized outcomes diverge significantly and *systematically* from expected ones. However, random errors must average out to zero in the long run and, hence, they cannot be used for 'learning' purposes.[6] Another way of putting this is that, if an economic agent makes a random error in a period and then makes different ones in subsequent periods in the wake of new shocks, *she will not be able to discern a systematic pattern in error-making* and will thus be unable to revise expectations accordingly. Consequently, error-elimination cannot proceed if errors are random. Now, if the environment changes in a way that the 'true' model also changes economic agents will need to revise their model. Otherwise, their errors could not remain random. According to PTKL such `revision´ can only be the result of a process of trial and error-elimination. But, as we argued above, if such a process is to come about errors must be systematic. Yet, the latter is precluded by REH. To conclude, REH lacks a truly learning mechanism grounded upon a trial and error-elimination process as required by PTKL. The only mechanism advocates of REH can rely on to justify the assumption that economic agents know the 'true' model is a simple extrapolation into the future of data obtained from the past. Yet, if this is to take place the stochastic environment needs to be assumed to be stochastically stable or ergodic. This explains why Lucas (1977) rightly suggests that REH only applies to situations of 'risk'.

Consequently, if REH is to be internally consistent its advocates need to assume that: (i) market participants do not need to learn the 'true' model, they simply know it, and (ii) the surrounding environment is ergodic so that the 'true' model does not change throughout. In an ergodic environment the only source of fallibility is economic agents´ inability to forecast future 'exogenous' random shocks.

3.2. Can our knowledge become 'optimal'?

Models that adopt REH assume that economic agents´ knowledge is perfectly adapted to the surrounding environment so that, as we noted above, the only source of fallibility economic agents exhibit is their inability to fully anticipate the 'exogenous' random shocks that will hit the economy. By contrast, PTKL implies that the learning process market participants go through *never* results in an optimal adaptation to the surrounding environment (Popper, 1994, p. 4). In particular, Popper emphasizes that the process of adaptation to the latter is often successful and often unsuccessful owing to the fact that some errors will inevitably *escape* and this possibility is one of the reasons why we are fallible (Popper, 1990, p. 47). Therefore, and due to the `less than complete´ elimination of errors, the adaptation of our knowledge is always *imperfect*. It follows that observed states of adaptation of knowledge cannot be the result of convergence to an optimum. Should *all* errors be purged, the process of adaptation to the surrounding environment would be perfect. At that stage, our knowledge would be fallible *only* to the extent that changes in the surrounding environment cannot be fully anticipated. As we noted above, this is the scenario implied by REH. By contrast, PTKL has it that our errors may well be systematic.

3.3. What, if the future is objectively open?

Lucas (1977, p. 15) argues that Muthian rationality 'will most likely be useful in situations in which the probabilities of interest concern a well defined *recurrent* event, situations of 'risk'. The latter correspond to environments in which individuals can infer the empirical regularities observed in the past and extrapolate them into the future. We argued above that the 'learning' process Lucas (1986) apparently has in mind makes sense solely in situations where the environment is ergodic. The reason is that, as we explained above, REH is logically inconsistent in a non-ergodic environment insofar as it does not possess a learning mechanism. Now, the ergodicity assumption logically implies that the actions undertaken by economic agents during the 'learning' process both before and after expectations become 'rational' do not affect the RE equilibrium. But this, in turn, implies that the latter exists *prior* to the beginning of the 'learning' process, i.e., it is pre-determined. In short, REH implies a deterministic world in which the future is a mere statistical reflection of the past. However, in a deterministic world human action is unduly constrained. Specifically, only actions that bring about *routine* change are allowed for. Thus, by stating that REH only applies to situations of 'risk' (or 'small worlds') Lucas (1977) restricts its validity to a deterministic and, hence, ergodic world. In such a world, there is very little room for fallibility. By contrast, Popper posits that the world is 'non-deterministic'; it is a world where the future is yet to be built as a result of our actions. It is only in a macro-theory in which it is *assumed* that the world is both non-deterministic and not pre-determined that economists can take due account of human fallibility.

4. A tentative conceptual framework for macroeconomics

We believe that PTKL provides macroeconomics with a conceptual framework which can potentially encompass a wide range of traditions in economics. Our starting point is the idea that agents´ decisions require the previous formation of expectations. Let

us think of the relation between expectations and *ex-post* realizations as a *negative* feedback mechanism whereby economic agents systematically revise their expectations in the wake of observed discrepancies between expectations and realized outcome. In principle, if the object of knowledge about which economic agents form expectations is stochastically stable or ergodic and the feedback mechanism is fast and accurate enough then the systematic operation of a trial and error-elimination mechanism should allow expectations to converge to equilibrium. As we suggested above, this appears to be the scenario contemplated in Lucas (1986).

Now, PTKL envisages an evolutionary process driven mainly by the interaction of two different feedback mechanisms. The first mechanism relates economic agents´ expectations and *ex-post* realizations. As advanced above, it is a negative feedback that allows agents to revise their expectations in the face of observed discrepancies between the former and the latter. An implication of PTKL is that this feedback mechanism is *imperfect* as some errors manage to escape so that, even if the surrounding environment is ergodic expectations do not eventually converge to an optimum. The second feedback mechanism stems from the impact on the object of knowledge (e.g., a market economy) of market participants´ beliefs, decisions, and actions. Popper posits it in his *Poverty of Historicism* (Popper, 1944, p. 89) and denotes it as the 'Oedipus effect' (hereafter OE) after Sophocles´ *Oedipus the King* tragedy. Unlike the previous feedback mechanism, its sign is *a priori* is ambiguous. If *negative,* then economic agents´ decisions will bring about changes in the surrounding environment that tend to narrow the gap between *ex-ante* expectations and *ex-post* realizations. This case corresponds to the scenario known in the economics literature as *self-fulfilling prophecies*.[7] There are many phenomena in economics which can be characterized by the presence of a negative OE such as: bank panics, speculative bubbles, sovereign debt crisis in countries belonging to a currency

union, etc. For instance, if deposit holders *believe* that their bank may go illiquid, they will try to draw their deposits as soon as possible irrespective of what the real liquidity situation of the bank is. In turn, by doing so, they will actually drive the bank illiquid. Through this mechanism, the gap between economic agents´ expectations and *ex-post* realizations tends to narrow and may eventually disappear. Situations characterized by the presence of a negative OE may lead to instability in the economy and, hence, some sort of institutional device may need to be set in place so as to short-circuit the former and prevent instability from actually taking place (e.g., deposit insurance system, debt monetization, etc.). By contrast, if OE is *positive*, economic agents´ beliefs, decisions, and actions will cause changes in the surrounding environment that tend to *widen* the gap between *ex-ante* expectations and *ex-post* realizations. Situations characterized by a positive OE are uncommon in macroeconomics.[8]

Soros (2013) alludes to a feedback mechanism equivalent to OE he denotes as the 'principle of reflexivity' and places it together with the 'principle of fallibility' at the centre of a new conceptual framework allegedly aimed at replacing current thinking in economics based on both REH and the Efficient Market Hypothesis. As Soros (2013) admits, the conceptual framework he proposes is inspired by Popper´s philosophy and, hence, it exhibits many similarities with the one being proposed here. However, and to the best of our knowledge, he does not associate his `principle of reflexivity´ with Popper´s OE. Nevertheless, we think there are some differences between the framework proposed by Soros and ours. More specifically, Soros (2013, p. 323) argues that REH represents a *limiting* case within his conceptual framework. According to him, REH can be characterized by the twin facts that (i) market participants´ revision of expectations can go forever, and (ii) the environment does not undergo significant changes. By contrast, we think that REH is incompatible with PTKL and the conceptual framework

that stems from it owing to the following facts: (i) REH presupposes that there is a pre-determined reality (i.e., the world is ergodic), (ii) REH lacks a `learning´ mechanism that allows market participants to ever get to know the 'true' model (i.e., agents´ errors are assumed to be random), and (iii) REH presupposes that `true´ knowledge can be attained (i.e., there is an `inductive logic´). In the rest of this section we explore in more detail some features and implications of the conceptual framework sketched above.

4.1. Ephemeral equilibrium

Next, as we argued above, PTKL implies that market participants´ decisions and actions interact with the surrounding environment by virtue of two interrelated feedback mechanisms. In turn, the interaction of these two feedback mechanisms will, by itself, generate endogenous change and, in the process, it will tend to make equilibrium fragile and *short-lived*.[9] This is because, as market participants revise their expectations in the aftermath of observed mistakes, the operation of the *negative* OE alluded to above may lead to changes in the previous equilibrium by bringing about *irreversible* changes in the economy. No wonder, the existence of OE highlights the limitations of the notion of equilibrium in macroeconomics and suggests that the study of *processes* of change may be more illuminating for policy purposes than the analysis of ephemeral equilibria. For instance, in the context of the NNS model, the level of aggregate demand is assumed to adjust passively to the (equilibrium) supply-side determined level of economic activity in the long run whereas in the framework we posit here the existence of a negative OE suggests that changes in the demand side of the economy will feed back into the supply-side (e.g., if the unemployment rate exhibits hysteresis). The picture that emerges is thus one where the demand and supply side of the economy influence each other so the level of economic activity can be said to be neither supply-side nor demand-side determined.

This idea is formalized algebraically in Shaikh (2013) who constructs a system of three differential equations capturing the dynamics of the actual, expected, and fundamental value of a certain macro-variable. He shows that a reduced two-dimensional dynamical system capturing the dynamics of the actual and expected values of the variable in question is stable in the sense that both values fluctuate around the fundamental value although, crucially, the latter may drift away. In other words, the economic system does not possess an (fixed) equilibrium to which the actual, expected, and fundamental value converge but, rather, it constitutes a 'moving centre of gravity'. The main prediction that stems from Shaikh´s model is that the economy will exhibit boom-bust cycles of varying amplitude and length.

4.2. Endogenous versus exogenous instability

Let us start by saying that the adoption of REH in macroeconomics *precludes* the exploration of the hypothesis that important malfunctions of market economies are the outcome of *systematic* errors in the formation of economic agents´ expectations. If REH is adopted, phenomena such as financial crisis or aggregate output fluctuations can only be ascribed to the occurrence of 'exogenous' random shocks. Let us dwell on this point. When assuming that economic agents do not make systematic mistakes advocates of REH are also compelled to assume that the economy is hit by 'exogenous' shocks. If agents´ expectations are unbiased, shocks can only be random and 'exogenous' insofar as agents can fully anticipate any trouble lying ahead stemming from endogenous forces and act consequently. Further, *REH implies that agents' mistakes always occur after the shocks have hit the economy in the sense that they are 'caused' by them*. Now, if we put these two elements together we have that REH implies that agents *never* make mistakes other than those ones induced by random and 'exogenous' shocks and, consequently, non-

random mistakes cannot be a source of instability. For instance, aggregate output fluctuations can solely be ascribed to factors 'external' to the economy such as abrupt political changes, 'natural' disasters, or technological revolutions. In short, REH biases macroeconomics by ruling out *by default* the possibility that macroeconomic instability is endogenously generated and, by doing so, it restricts itself to a 'small world' in the sense of Savage (1954).[10]

4.3. Evolutionary macroeconomics

As we have noted above, the occurrence of random mistakes in the wake of 'exogenous' random shocks is the only source of fallibility allowed for in models that adopt REH. However, this is a severely restricted approach to fallibility. To be sure, accounting fully for fallibility implies recognition that economic agents´ mistakes may be *systematic* owing to the fact that: (i) we can never 'prove' that a given hypothesis is 'true', (ii) learning and, hence, knowledge is always imperfect, and (iii) the surrounding environment undergoes unpredictable changes brought about (endogenously) by agents´ beliefs, decisions, and actions. As a result of it, agents´ will typically face situations of 'uncertainty' in the sense of Knight (1971[1921]) and Keynes (1920) or, equivalently, they will find themselves in 'large worlds' (Savage, 1954). In such scenario, fallibility can be modelled by the operation of the two feedback mechanisms described above.

Now, the interaction of the two feedback mechanisms posited above may bring about endogenous change in a way that is strongly reminiscent of evolutionary systems. An example of the latter is Holland´s (1992) concept of 'Complex Adaptive System' or (CAS). If we think of a market economy as a particular type of CAS then its evolution can be modelled as driven by the interaction of *credit assignment* and *rule discovery* procedures. Both procedures constitute complex feedback systems which, on the one

hand, reinforce (discourage) the use of rules that have performed well (badly) in the past (‘credit assignment procedure’) and, on the other hand, produce new rules by combining the rules that have performed well in the past. The implementation of new rules will be aided by the formulation of ‘internal models’ that help anticipating the consequences of the adoption of certain rules. Importantly, and regardless of whether anticipations prove to be right or wrong, once such `internal models´ are undertaken, they will bring about *irreversible* changes in the surrounding environment thereby forcing CAS to re-evaluate its own ‘internal models’. Thus, the process of adaptation is never-ending and cannot be optimal. As noted in Beinhocker (2013), forecasting is extremely difficult in a CAS yet this does not prevent it from generating `predictions´, the latter being understood as the deductive logical consequences of a theory. An example of this is evolutionary biology.

The example of CAS reveals that, by adopting REH, macroeconomics becomes incompatible with an evolutionary approach. First, as we argued above, REH is not a hypothesis but an assumption. By contrast, according to PTKL knowledge constitutes the *product* of an evolutionary process whereby false conjectures tend to be eliminated over time. Second, many of the changes that take place in the economy in the wake of decisions and actions adopted by market participants may be ‘irreversible’ in the sense that, once they are undertaken, it is extremely costly to undo them or they simply cannot be undone because the surrounding environment may have changed substantially as a result of the decisions made. However, if some changes are ‘irreversible’ the economy will exhibit *path-dependency*. To be sure, the latter implies non-ergodicity and this, in turn, precludes REH.

We believe that if contemporary macroeconomics is to fully account for human fallibility it will need to adopt an *evolutionary* approach. Several traditions in economic analysis have adopted such an approach in the past, some well-known examples being the

(Old) Institutionalist School of Veblen, Mitchell and Commons, the Austrian School of Schumpeter, Menger, and Hayek, the Evolutionary school (Nelson & Winter, 2002), Complexity Economics (Arthur, 2013), Imperfect Knowledge Economics (Frydman & Goldberg, 2007), and those Post-Keynesians who have adopted Minsky's interpretation of Keynes's *General Theory* (Minsky, 1975). It is beyond the scope of this essay to evaluate whether and to what extent these approaches are fully compatible with PTKL. However, let us note that a common feature of all of them is that change is generated *endogenously* by a negative feedback mechanism connecting observed discrepancies between expected and realized outcomes, elimination of errors, and generation of new (tentative) conjectures.

4.4. The partial autonomy of expectations and its implications

In the context of the NNS model it is assumed that in the presence of (full) price and wage flexibility a market economy will reach steady-state equilibrium. However, its advocates commonly argue that in the presence of nominal and real rigidities aggregate output may deviate *temporarily* from steady-state equilibrium. In particular, this will be the case if the economy is hit by 'exogenous' shocks. Policy analysis in the NSS model thus consists of determining the optimal values of policy instruments given a fully pre-specified macroeconomic model where economic agents know both the 'true' model of the economy and the stochastic properties of 'exogenous' shocks. Now, the NNS model implies that market participants´ expectations do not play an *autonomous* role in the determination of the values of the variables of the model. This is captured in Frydman and Goldberg (2013, p. 143) who note that 'REH, by design, imposes exact consistency between the sharp prediction — a single probability distribution of outcomes — implied by an economist´s aggregate model and the probability distribution representing

participants´ forecasting strategies'. In order words, REH implies that economic agents´ expectations adapt *passively* to the forecasts generated by the model itself. This way, macroeconomists may *assume* (rather than prove) the existence of a stable equilibrium where: (i) expectations are always and fully realized, and (ii) there are no endogenous forces pushing the economy away from it. By contrast, the possibility that market participants´ decisions change the surrounding environment implies that expectations do play, at least partly, an *autonomous* role in the sense that they do not adapt passively to a pre-determined equilibrium. No wonder, the recognition of the partial autonomy of expectations takes us back to Keynes´s *General Theory* where he makes it clear that:

'The considerations upon which expectations of prospective yields are based are partly existing facts which we can assume to be known more or less for certain, and partly future events which can only be forecasted with more or less confidence... We may sum up the state of psychological expectation which covers the latter as being the *state of long-term expectation*' (Keynes, 1936, pp. 147-148).

Keynes (1936, ch. 12) argues that the 'state of long-term expectations' depends partly on the 'state of confidence' with which we make the most probable forecast we can make. In turn, the former affects the marginal efficiency of capital and liquidity preference and, through the multiplier effect on consumption demand, it determines the level of economic activity. The partial autonomy of expectations hinders a theoreticians´ attempt to obtain an overarching account of the economy and, arguably, constitutes a key difference between the natural and the social sciences. As argued in Hayek (1943), it is only in the social sciences that our *interpretation* of a situation no matter whether it is right or wrong becomes an integral part of the situation thereby affecting subsequent

developments. Further, and to the extent that we understand the surrounding world via the 'internal models' we create for that purpose, our (partial) understanding of the world will affect our decisions and, in this way, it may affect the world itself. Acceptingly, the partial autonomy of economic agents´ expectations with respect to the current structure of the economy poses a formidable methodological challenge to macroeconomists. The challenge is to develop a conceptual framework that takes account of the simultaneous evolution of the economy´s structure and economic agents´ understanding of the latter.

4.5. Rationality

The notion of rationality embedded in REH is contingent on the crucial premise that the 'true' model of the economy is either known or knowable. This is implicit in Muth´s (1961) description of REH as a purely descriptive hypothesis and in Lucas´ reinterpretation of REH as an assumption about how every market participant forecasts. As Frydman and Goldberg (2013) note, Lucas took for granted that market participants can construct a fully predetermined model that provides an account of how the economy evolves over time and, consequently, he argued that it would be 'irrational' for them to make systematic forecast errors since this would reveal the presence of unexploited profit opportunities (Lucas, 2001, p.13). Therefore, even if we disregard the conceptual difficulties that arise if an attempt is made to show REH as the end-result of an *eductive* process the question emerges what the appropriate benchmark for individual rationality is when the 'true' model of the economy is unknown and unknowable.[11] We believe the answer is in PTKL. According to it, 'rational' behaviour consists of *making conjectures and the corresponding decisions and, if necessary, revising them in the face of observed discrepancies between expected and realized outcomes*. This notion of rationality does not imply that rational behaviour is always or often *effective*. Rather, all that is required

from economic agents is that they revise their decisions in the light of their observed mistakes. In particular, a decision which initially appears to be correct *a priori* given the information available to decision-makers may turn out to be wrong *ex-post* for reasons unrelated to the reasonableness of the decision previously made such as, for instance, the occurrence of unforeseen changes in the surrounding environment.

Next, Popper´s notion of individual rationality (PR) bears some resemblance to Simon´s (1976) notion of 'bounded' rationality (BR) in that, unlike the accompanying notion of 'substantive' rationality (SR), rational behaviour is not assumed to be optimal. Nevertheless, as we show below, it exhibits several differences with it.[12] Table 1 below illustrates the key differences and similarities between PR and the notion of SR and BR coined by Simon (1976, 1979). First, SR presumes that the 'true' model of the economy is known or knowable by economic agents whereas BR presumes that the former is not known for practical purposes but could be potentially known if the costs of searching for information were low enough and the computational difficulties faced by economic agents were not insurmountable. This is apparent, for instance, in Simon (1979, p. 502) when he makes it clear that BR is 'a residual category — rationality is bounded when it falls short of omniscience'. As Simon (*op. cit*.) explains, there are three different types of failures of omniscience: (i) failures of knowing all the alternatives, (ii) uncertainty about exogenous events, and (iii) inability (due to computational capacity constraints) to calculate the full consequences of every possible choice. Further, he argues that if the alternatives for choice are not given initially market participants will need to *search* for them and, hence, a theory of BR must incorporate a 'theory of search'. However, the implicit premise is that if search costs were low enough and computational constraints were negligible market participants could know all the possible alternatives for choice. In other words, Simon apparently takes for granted that there exists an 'inductive logic' so

that what prevents economic agents from knowing the 'true' model of the economy is the presence of both high information search costs and computational constraints.[13] In contrast to this, PR presupposes that the 'true' model of the economy is neither known nor knowable, the main reason being the conjectural nature of knowledge.

<table>
<tr><th>Approaches</th><th>Domain</th><th colspan="2">Dichotomy</th></tr>
<tr><td rowspan="3">Conventional approaches to rationality: The 'true' model is either known or knowable</td><td rowspan="2">'Substantive' rationality: it
applies only to situations of 'risk'</td><td>Rational behaviour</td><td>Optimal (attainable)</td></tr>
<tr><td>Irrational behaviour</td><td>Sub-optimal</td></tr>
<tr><td rowspan="2">'Bounded' rationality: it applies both to situations of 'risk' & 'Knightian' uncertainty</td><td rowspan="2">Behaviour is generally `rational´
(e.g. heuristics)</td><td>Optimal
(Unattainable)</td></tr>
<tr><td>Behaviour is generally sub-optimal</td></tr>
<tr><td>Popper´s approach to rationality
(The 'true' model is neither known nor knowable)</td><td>It applies both to situations of 'risk' & 'Knightian' uncertainty</td><td>Behaviour is generally `rational´(i.e., corrigible)</td><td>No dichotomy</td></tr>
</table>

Table 3: Popper´s notion of rationality and alternative approaches to rationality in economics

Now, as Lucas (1977) insists, 'in cases of [Knightian] uncertainty, economic reasoning will be of no value' which entails that SR and REH only apply to situations of 'risk' or 'small worlds'. By contrast, BR and PR apply both to situations of 'risk' and 'Knightian' uncertainty. In situations of 'risk' BR basically consists of using rules (e.g. algorithms) to estimate the probabilities of an exhaustive list of potential outcomes by extrapolating information obtained from the past. Similarly, in situations of 'risk' PR

amounts to re-estimating periodically the numerical probabilities of a list of potential outcomes in the light of observed discrepancies between estimated probabilities and realized relative frequencies. However, as Lucas (1977) argues, SR does not apply in situations characterized by 'Knightian' uncertainty due to the impossibility of applying (neoclassical) economic theory. By contrast, both BR and PR apply in this case thereby leaving some room for 'economic reasoning' even when the surrounding environment is such as to preclude the assignment of numerical probabilities to each possible outcome. There is, by now, a large literature on rational behaviour under 'Knightian' uncertainty where the former consists of the use of different heuristics (Gigerenzer & Gaissmaier, 2011) and the adoption of social rules (Hodgson, 1997). We believe these approaches to rational behaviour are compatible with PKTL.

Next, both SR and BR imply an undesirable dichotomy between optimality and sub-optimality. The dichotomy is *explicit* in SR where `rational´ behaviour is associated to optimality whereas `irrational´ behaviour is synonymous with sub-optimal behaviour (Becker, 1962). By contrast, we believe the dichotomy is *implicit* in BR where optimal behaviour is presented as being generally unattainable but `potentially attainable´ under favourable (albeit very unusual) circumstances. In particular, BR implies that results will generally be sub-optimal due to the shortcomings of market participants´ cognitive abilities which prevent them from processing, storing, and analysing a large quantity of information. Yet, this implies that optimal behaviour constitutes a benchmark against which the actual degree of sub-optimality of economic agents´ behaviour can be *a priori* determined. By contrast, such dichotomy does not exist in the case of PR where there is no such thing as an 'optimal' benchmark. Rather, PR implies that market participants´ choices will never be optimal (except by sheer chance) *even* if their cognitive abilities and computational power are unbounded insofar as knowledge — including knowledge

about the 'true' model of the economy — is always *conjectural*. In other words, the notion of optimal choice that pervades contemporary macroeconomics makes sense only if the 'true' model of the economy is either known or knowable by market participants. By contrast, PTKL does not presuppose that `true´ knowledge is attainable; it implies that economic agents´ behaviour is normally rational insofar as Popper´s 'principle of transference' (Popper, 1972, p. 6) posits that `what is true in philosophy is also true in psychology´. In short, the notion of 'rationality' that stems from PTKL is inextricably linked to the *fallibility* of human knowledge.

5. Summary and conclusions

Macroeconomic theory has been dominated over the last decades by REH which implies that economic agents are assumed to know the 'true' model of the economy so that their expectations are unbiased. The assumption that economic agents do not make systematic mistakes implies, in turn, that economies are assumed to be inherently stable and that macroeconomic volatility is caused exclusively by 'exogenous' random shocks. REH is a core element of the NNS model which has played a dominant role in theory and policy over (at least) the last decade and which, according to some commentators, bears some responsibility for the generation of the last financial crisis. We ascribed the difficulties of REH and the NNS model to their flawed epistemological and ontological foundations. Specifically, we argued that the adoption of REH biases macroeconomics and reduces its practical relevance by restricting its applicability to situations of 'risk' (Knight, 1971[1921]) or 'small worlds' (Savage, 1954). In such contexts, the solely source of fallibility is the inability by economic agents to anticipate 'exogenous' random shocks. We argued that this is too constraining a conception of human fallibility as it ignores fundamental sources of the latter such as the logical impossibility of knowing the

'true' model of the economy owing to the lack of an 'inductive logic', the imperfect nature of learning, and the changing nature of the environment in which economic agents make economic decisions. We went on to claim that PTKL provides macro-theory with solid epistemological and ontological foundations in that it enables it to take *full* account of human fallibility.

If macroeconomics is to reorient itself to take full account of human fallibility it needs to focus on the changes in the surrounding environment brought about by market participants´ beliefs, decisions, and actions. In particular, we argued that, in the context of a market economy, change and evolution can be viewed as generated *endogenously* by the interaction of the two feedback mechanisms: (i) a negative feedback mechanism whereby market participants revise their previous expectations in the light of observed systematic discrepancies between expected and realized outcomes, and (ii) a feedback mechanism of ambiguous sign whereby market participants´ decisions and actions may alter the surrounding environment in an unpredictable way. We then argued that the endogeneity of change and evolution in market economies generates two predictions: (i) macroeconomic equilibrium is fragile and, hence, *short-lived* thereby making it difficult for theorists and policy-makers to exploit it, and (ii) market economies exhibit boom-bust cycles of varying amplitude and duration so that there is *never* a repetition of the same cycle. The former suggests that the importance of the notion of equilibrium in the context of macroeconomics tends to be overemphasized and that more attention needs to be paid to the causes and features of processes of transformation over time instead of characterizing (stable) equilibrium. Finally, we argued that, unlike conventional notions of rationality in economics such as Simon´s twin notions of 'substantive' and 'bounded' rationality, the notion of rationality associated to PTKL does not generate a *dichotomy*

between `optimal´ and `sub-optimal´ behaviour since, unlike them, PTKL presupposes the inexistence of an `inductive logic´.

Lastly, we noted that some of the concepts we associated to PTKL can be also identified in other approaches in economics that have developed outside mainstream economics which, according to us, suggests that the development of a new conceptual framework for macroeconomics would benefit substantially from a re-examination of these approaches. Further, we believe there are other branches of knowledge economists can also take advantage of and whose conceptual frameworks are *a priori* compatible with PTKL. Some examples of this are: evolutionary psychology, evolutionary biology, and 'complex adaptive systems' theory. In short, we believe the development of a new macro-theory along PTKL will require that economists adopt a more interdisciplinary and less pretentious approach than has been the case in the recent past. To paraphrase Popper 'the future is objectively open'. Thus, it is in our own hands to change it. Let us hope that economists take up the challenge.

References

Arrow, K. J. (1978) The Future and the Present in Economic Life. *Economic Inquiry*, 16 (2):157-69.

Arthur, W. B. (2013) Complexity economics: a different framework for economic thought. *Institute for New Economic Thinking Research*. Note #033, March.

Becker, G. (1962) Irrational behavior and economic theory, *Journal of Political Economy*, 70(1), pp. 1-13.

Beinhocker, E. D. (2013) Reflexivity, complexity, and the nature of social science. *Journal of Economic Methodology*, 20 (4), 330-342.

Boland, L. A. (2003) *The Foundations of Economic Method: A Popperian Perspective,* 2nd edition. London: Routledge.

Cross, R. (1982) The Duhem-Quine Thesis, Lakatos and the Appraisal of Theories in Macroeconomics, *Economic Journal*, 92(366), pp. 320-340.

Davidson, P. (1991) Is Probability Theory Relevant for Uncertainty? A Post Keynesian Perspective. *Journal of Economic Perspectives*, 5 (1), 129-143.

Davidson, P. (1996) Reality and Economic Theory, *Journal of Post Keynesian Economics*, 18(4), pp. 479-508.

De Grauwe, P. (2009) Top-down versus bottom-up macroeconomics. Paper presented at the conference "What is wrong with modern macroeconomics". CESifo, Munich, 6-7 November 2009.

Farmer, R. E. (2013) Animal Spirits, Persistent Unemployment, and the Belief Function. In R. Frydman and E. S. Phelps (eds.), *Rethinking Expectations: The Way Forward for Macroeconomics*, 251-276, Princeton: Princeton University Pres.

Frydman, R. & Goldberg, M. D. (2007) *Imperfect Knowledge Economics.* Princeton: Princeton University Press.

Frydman, R. & Goldberg, M. D. (2011) *Beyond Mechanical Markets: Asset Price Swings, Risk, and the Role of the State*. Princeton: Princeton University Press.

Frydman, R. & Goldberg, M. D. (2013) The Imperfect Knowledge Imperative in Modern Macroeconomics and Finance Theory. In R. Frydman and E. S. Phelps (eds.), *Rethinking Expectations: The Way Forward for Macroeconomics*, 130-165, Princeton: Princeton University Press.

Frydman & Phelps (2013) Which way forward for macroeconomics and policy analysis? In R. Frydman and E. S. Phelps (eds.), *Rethinking Expectations: The Way Forward for Macroeconomics, 1-48.* Princeton: Princeton University Press.

Gigerenzer, G. & Gaissmaier, W. (2011) "Heuristic Decision Making." *Annual Review of Psychology* (62): 451-482.

Goodfriend, M. & king, M. (1997) The New Neoclassical Synthesis and the Role of Monetary Policy. *NBER Macroeconomics Annual*, 231-283.

Guesnerie, R. (2013) Expectational Coordination Failures and Market Volatility. In R. Frydman and E. S. Phelps (eds.), *Rethinking Expectations: The Way Forward for Macroeconomics*, 49-67, Princeton: Princeton University Pres.

Hayek, F. A. (1948[1937]) Economics and Knowledge. *Individualism and economic order*, 33-56, Chicago (IL): University of Chicago Press.

Hayek, F. A. (1943) The Facts of the Social Sciences. *Ethics*, 54(1), 1-13.

Hodgson, G. M. (1997) The ubiquity of habits and rules. *Cambridge Journal of Economics*, 21 (6), 663-684.

Holland, J. H. (1992) Complex Adaptive Systems. *Daedalus — Journal of the American Academy of Arts and Sciences*, 121 (1), 17-30.

Hume, D. (2006[1748]) *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press; New York: Oxford University Press, 2006

Keynes, J. M. (1920) *A Treatise on Probability*. London: Macmillan.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. London: Macmillan.

Knight, F. H. (1971[1921]) *Risk, Uncertainty, and Profit.* Chicago: The University of Chicago Press.

Knudsen, C. (1993) Equilibrium, perfect rationality and the problem of self-reference in economics. In Mäki, U., Gustafsson, B., and Knudsen, C. (eds.), *Rationality, Institutions and Economic Methodology*: 133-70. London: Routledge.

Krugman, P. (2009) How Did Economists Get It So Wrong? New York Times. http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html?pagewanted=alland_r=0 Accessed December 3rd.

Lucas, R. E. (1977) Understanding Business Cycles. In Carnegie-Rochester Conference Series on Public Policy (ed.), Vol. 5. Stabilization of the Domestic and International Economy: 7-29.

Lucas, R. E. (1986) Adaptive behaviour and economic theory. In Brunner, K., Meltzer, A.H., R. M. Hogarth and M. W. Reder (eds.), *Rational Choice: The Contrast between Economics and Psychology*. Chicago and London: University of Chicago Press: 217-242.

Lucas, R. E. (2001) Professional Memoir. Available at: http://home.chicago.edu.

Merton, R. K. (1948) The self-fulfilling prophecy. *Antioch Review*, 8 (2), 193-210.

Minsky, H. P. (1975) *John Maynard Keynes*. New York: Columbia University Press.

Muth, J. F. (1961) Rational Expectations and the Theory of Price Movements. *Econometrica*, 29 (3): 315-335.

Nelson, R. R. & Winter, S. G. (2002) Evolutionary Theorizing in Economics. *Journal of Economic Perspectives*. 16 (2): 23-46.

Phelps, E. S. (2007) Foreword. In Frydman, R. and Goldberg, M. D. (eds.), *Imperfect Knowledge Economics*, pp. xiii-xx, Princeton: Princeton University Press.

Popper, K. R. (1944) The Poverty of Historicism I. *Economica*, 11 (42), 86-103.

Popper, K. R. (1972) *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

Popper, K. R. (1990) *A World of Propensities*. Bristol (UK): Thoemmes.

Popper, K. R. (1994) *The Myth of the Framework: In defence of science and rationality*. London: Routledge.

Rosenberg, A. (2013) Reflexivity, uncertainty and the unity of science. *Journal of Economic Methodology*, 20 (4), 429-438.

Savage, L. (1954). *The Foundations of Statistics*. New York: Wiley.

Shaikh, A. (2013). On the role of reflexivity in economic analysis. *Journal of Economic Methodology*, 20 (4), 439-445.

Simon, H. A. (1976). From Substantive to Procedural Rationality. In Spiro J. Latsis (ed.), *Method and Appraisal in Economics*, 129-148. Cambridge: Cambridge University Press.

Simon, H.A. (1979). Rational Decision Making in Business Organizations. *American Economic Review*, 69 (4), 493-513.

Soros, G. (2013). Fallibility, reflexivity, and the human uncertainty principle. *Journal of Economic Methodology*, 20 (4), 309-329.

Stiglitz, J. E. (2011). Rethinking Macroeconomics: What Failed, and How to Repair it. *Journal of the European Economic Association*. 9 (4): 591-645.

Tobin, J. (1981). The Monetarist Counterrevolution Today — An Appraisal. Economic Journal, 91 (361): 29-42.

Woodford, M. (2009). Convergence in Macroeconomics: Elements of the New Synthesis. *American Economic Journal*: Macroeconomics. 1 (1): 267-279.

Notes

[1] For instance, Frydman & Phelps (2013, p. 6) argue that 'in a leap of faith that transformed macroeconomics and finance for generations, Lucas presumed that the right theory of capitalist economies, which arguably thrive on nonroutine change, is a fully predetermined model that assumes that such change is unimportant'.

[2] However, Muth (1961, p. 317) notes that, for the purpose of analysis, he focuses on a 'specialized' or restrictive form of REH where it is assumed that: (i) the random disturbances that hit the economy are normally distributed, (ii) certainty equivalence holds for all the variables that are to be forecast, and (iii) the equations of the system are linear.

[3] Simon (1976, p. 130) identifies human behaviour as being *substantively* rational 'when it is appropriate to the achievement of given goals within the limits imposed by given conditions and constraints'.

[4] Specifically, if the stochastic process is ergodic, then for an infinite realization of it the time statistics will coincide with the space statistics. If the realization of the stochastic process is finite, time and space statistics will coincide except for the presence of random errors albeit they will tend to converge as the number of observations increases. Thus, if the surrounding environment is assumed to be ergodic, then statistics calculated from past time series or cross-sectional data will be reliable estimates of the space statistics that will occur at any future date (Davidson, 1996, p. 480).

[5] To be sure, refutation does not prove a theory false in the sense that negative empirical results are *never* definitive. In other words, no conclusive disproof of a hypothesis can ever be produced for it is always possible to say that the experimental results are not reliable. In particular, the empirical testing of any hypothesis is necessarily accompanied by a certain number of `auxiliary´ empirical hypotheses so that, in the wake of a negative result, the theoretician may be tempted to blame any of the `auxiliary´ hypotheses

rather than the `central´ hypothesis for the occurrence of the negative result. This problem is known in the literature as the `Duhem-Quine´ thesis. A discussion of the latter in the context of macroeconomics is in Cross (1982).

[6] Under REH forecast errors stem only from the occurrence of 'exogenous' random shocks since the model that generates forecasts is assumed to coincide with the 'true' model. In turn, random errors are assumed to be generated by a probability distribution of zero mean.

[7] According to Merton (1948, p. 195) a 'self-fulfilling prophecy' is an initially false definition, conception or interpretation of the situation that evokes a new behaviour which makes the originally false definition come true.

[8] Let us provide an example related to social psychology. Let us assume that there is a widespread belief in the media and the sports community that a certain team is to win a championship. Now, on the one hand, the members of the team may well relax since everybody takes for granted that they will win. On the other hand, rival teams are likely to work and train harder since they are aware that they will only have a chance to win if they improve their performance substantially. The joint effect of the initial belief on both the winner-to-be team and its weaker rivals may end up — through the operation of a *positive* OE — making one of the rival teams eventually win the championship contrary to the initial expectations. In other words, the operation of a *positive* OE in this case leads to a widening of the gap between agents´ expectations and realized outcomes.

[9] In this respect, Rosenberg (2013) draws a useful distinction between `biological´ and `cultural´ domains. According to him, the difference in the nature of equilibrium in these two domains is only *one of degree*. As he explains, in the biological domain the interaction of different feedback mechanisms produces stable cycles that exhibit a high degree of persistence owing to the fact that the former are 'held in check by environments that change with geological slowness' (2013, p. 435). By contrast, he argues that the typical environment in cultural domains is subject to very rapid change. In particular, their lifetimes are inversely related to the rate of change of the cultural environment itself. Consequently, equilibrium in such domains is *short-lived*.

[10] The dichotomy between exogenous and endogenously-created instability is also captured in De Grauwe (2009) who distinguishes between 'top-down' and 'bottom-up' systems. The former are systems where one or more agents fully understand the system as in the models adopting REH. By contrast, the latter are systems in which no individual understands the whole picture so that rationality consists of a trial-and-error selection process. De Grauwe (2009) recognises that the latter constitutes an example of rational behaviour which may result in a business cycle with a large *endogenous* component. A similar remark is in Arthur (2013, p. 3). However, neither De Grauwe (2009) nor Arthur (2013) mention that the notion of a trial and error-elimination process as the characterization of rationality has its philosophical foundation in PTKL.

[11] As Guesnerie (2013, p. 52) points out, the RE equilibrium is a *Nash equilibrium*. Yet, the possibility of attaining an equilibrium through an `eductive´ process — a process that takes place in people´s minds whereby equilibrium is achieved through *systematic reasoning* of the decision-makers involved — faces insurmountable conceptual difficulties as illustrated in Knudsen (1993).

[12] According to Simon (1979), economic agents´ knowledge is subject to three types of constraints: (i) a limited ability to process, analyse, and store information, (ii) uncertainty, and (iii) the presence of social institutions. His notion of 'bounded' rationality stems from the fact that the existence of these constraints prevents economic agents from 'optimizing'.

[13] Or as Boland (2003, p. 40) puts it, Simon (1979) does not really deny Inductivism, only the feasibility of inductive knowledge.