

Ontological constraints on rational irrationality: the case of electoral preferences

Mihai UNGUREANU*

National University of Political Studies and Public Administration, Bucharest, Romania
mihai.ungureanu@snsa.ro

Abstract. *This paper is concerned with Caplan's (2000, 2001, 2006) rational irrationality model and its application on electoral preference formation. According to this model, individuals consume irrationality in a rational manner i.e. they maximize their total welfare through an exchange between their material and psychological welfare. Using the contributions of an important debate from the philosophy of economics regarding the imposition of ontological constraints on models, unrealistic assumptions and explanatory unification, two arguments are formulated: First, Caplan's model is ontologically bewildered – it contains opposed statements about the psychological descriptiveness of rational irrationality. Second, even though Caplan is apparently aware of the tension between ontological and tractability considerations he seems to privilege the latter.*

Keywords: rational irrationality, unrealistic assumptions, realism, ontological constraints, electoral preferences.

JEL Classification: P51.

REL Classification: 5J.

* Scholar within the project "Doctoral and Postdoctoral Fellowships for young researchers in the fields of Political, Administrative and Communication Sciences and Sociology" POSDRU/159/1.5/S/134650, financed through the Sectoral Operational Programme for Human Resources Development 2007-2013, co-financed by the European Social Fund.

1. Introduction

Caplan (2000, 2001a, 2006) formulated the rational irrationality model as the main vehicle of analyzing electoral preferences. Following the tradition started by Akerlof and Dickens (1989), the model introduces elements of irrationality in the neoclassical economics framework. Agents have preferences not only for available states of the world but also on their beliefs about those states of the world. As a result, Caplan argues, they maximize a combination of psychological welfare (from consuming irrationality) and material welfare. The methodological details of this model are a good argument for its selection for a thorough analysis of the ontological status of its unrealistic assumptions. For this purpose I focus on the relevance of the ontological constraints on defining local scientific realism (Mäki, 2005), building the Musgrave-Mäki-Hindriks typology of assumptions, defining models as surrogate systems (Mäki, 2009a) and on the ideal of explanatory unification. The debate started with Friedman's (1953) argument that a model should be judged only by the precision of its predictions regardless of its unrealistic assumptions. This position, later labeled as instrumentalist⁽¹⁾, generated a rich literature in the philosophy of economics containing the debate between realists and antirealists about the legitimacy of using unrealistic assumptions. It has been argued (Musgrave-Mäki-Hindriks) that the ontological status of assumptions is in fact relevant and the decision of their acceptability should be based on the functions they serve inside the models. Taking these statements as a starting point, this paper analyzes the ontological status of rational irrationality and formulates two important arguments: First Caplan's model is ontologically bewildered – it contains conflicting statements, some placing the model in Friedman's instrumentalist tradition, and others in a realist tradition, rather resembling the concerns of behavioral economics. Second, Caplan's choice of modeling 'belief change' in the utility maximization framework seems to be rather too methodologically constrained than necessary in order to unveil the underlying mechanism of the target world. Additionally (and related to the previous argument), Caplan insufficiently argued for the superiority of its rational irrationality model against its competitors in explaining observed behavior. These arguments are based on the main application⁽²⁾ of rational irrationality – electoral preference formation. The paper is structured in two parts. The first concerns the debate regarding scientific realism, the functions of unrealistic assumptions and their role in the attempts of explanatory unification. The second part presents and analyzes the rational irrationality model on the coordinates fixed in the first part.

2. Neoclassical ontology

A traditional attack against neoclassical economics was to accuse it of using unrealistic assumptions to model economic behavior.⁽³⁾ Friedman (1953) has provided economists with an important justification for using unrealistic assumptions. His argument is that the truth/realisticness of assumptions is irrelevant in accepting a theory, as long as its predictions are correct (Friedman, 1953: pp.8-9) Moreover, in Friedman's words: "the more significant the theory, the more unrealistic the assumptions [...] To be important, therefore, a hypothesis must be descriptively false in its assumptions" (Friedman, 1953: p. 14). To support this view, Friedman offers the example of Galileo's law of freely falling bodies ($s = 1/2gt^2$) and shows that one of its assumptions (i.e. the vacuum assumption) is certainly simplifying but this is not problematic per se: "bodies that fall in the actual atmosphere behave *as if* they were falling in a vacuum [...] The formula is accepted because it works, not because we live in an approximate vacuum"⁽⁴⁾ (Friedman, 1953: p. 18) Likewise, Friedman argues, the assumption of utility maximization (and other assumptions), although unrealistic, should be judged by the success of predictions derived with their help. This position has quickly become the standard justification of the simplifications used by economists. At the same time it received significant criticism regarding its lack of concern for the ontological status of the simplifications used in modeling.

2.1. Local realism, surrogate systems and explanatory unification in neoclassical economics

2.1a. Local realism and anti-realism

Friedman's position, but also its impact on economists produced substantial subsequent literature on the dispute between realists and instrumentalists. Friedman was labeled instrumentalist (Bear, Orr, 1967: pp. 192-193, Boland, 1979: p. 505, Caldwell, 1980: p. 366) and initiator of the formalist revolution in economics (Blaug, 2002: p. 30) and the ontology of its approach was often attacked as incompatible with a realist program of research. A classic definition of the realist tradition as "global doctrine about science" (Mäki, 2005: p. 231) is in terms of "mature and predictively successful scientific theories as well-confirmed and approximately true of the world."⁽⁵⁾ (Psillos, 2005: p. xvii). Realism also holds that "a full and just explication of theoretical discourse in science requires commitment to the existence of unobservable entities."⁽⁶⁾ (Psillos, 2005: p. 2) and that they are independent of our ability to collect direct evidence of their existence (Psillos, 2005: p. 44) resulting in postulating a mind-independent reality (i.e. from our sensory data or from the structure of our minds). Mäki (2009a: p. 74) argues, however, that some of the descriptions of scientific realism must be thought to be rather local. They vary from case to case – for example, references to

unobservable entities appear to be imposed by exclusive reference to one discipline i.e. Physics – (Mäki, 2005: p. 235) and therefore realism must be reformulated in more general terms. Such formulation would be consistent with its use in the social sciences, and thus with economics. Mäki argues that social reality cannot be mind-independent (Mäki, 2005: p. 243), and therefore a weaker condition should take its place – the science independence – the social world exists independently of theory – the social reality is not created by the social science (Mäki, 2005: p. 245). Also, social sciences do not generally use unobservable entities. They usually employ manifest images or commonsensibles (i.e. beliefs, norms, roles etc.). Moreover, social sciences should not be evaluated by their predictive success or by their maturity. All that is required is that they should be concerned about the ontological status of their theories (Mäki, 2005: p. 239). Hence, an anti-realist attitude would be one that is not concerned with the ontology of theories or models. In fact, as Mäki noted, realism is primarily an “ontological doctrine” (Mäki, 2005: p. 236) and it is to be expected that antirealism would not be concerned with the ontology of scientific theories.

2.1b. Ontological constraint on models and explanatory unification

In this view about realism, models must be understood as “representations of some target (such as a real world system, a set of data, or a theory” (Mäki, 2009a: p. 75). They are “surrogate systems [i.e. mediating vehicles in attempts to gain indirect epistemic access to the real world” (Mäki, 2009a: p. 77)], and economic research is directly concerned with the identification of the properties of these surrogate systems - rather than direct identification of the properties of real systems (Mäki, 2009a: p. 76). A realist intuition is that the model world and the target world should resemble each other, especially when considering that the use of unrealistic assumptions is unavoidable in the construction of economic models. These “strategic falsehoods” (Mäki, 2009a: p. 78) have the goal of “isolating some important dependency relation or causal factor or mechanism from the involvement and influence of the rest of the universe”⁽⁷⁾ (Mäki, 2009a: p. 78). The same vision about the ontological constraint is used concerning explanatory unification: “Much of explanatory activity in economics is driven by the ideal of *unification*: the urge to explain much by little, to explain many kinds of phenomena in terms of the same parsimonious explanatory principles” (Mäki, 2009a: p. 86). Mäki (2001b) distinguishes between two kinds of explanatory unification, ontological and derivational. “Ontological unification is based on the representational capabilities of theories, while derivational unification is based on their inferential capabilities” (Mäki, 2001b: p. 491). In other words, “ontological as opposed to mere derivational unification gives priority to entities rather than sentences; reference and representation rather than inference and derivation; discovery rather than imposition.” (Mäki, 2001b: p. 499) Mäki therefore

distinguishes between “unification as formal constraint” and unification as factual discovery” (Mäki, 2001b: p. 503). These ideas are exemplified by the use of homo economicus in neoclassical economics. On one hand, this assumption can be used as a methodological constraint which prescribes that the results must be derived with its help; on the other hand, when utility maximization is used as discovery, unification is actually a matter of determining how the world really works, and the principle of utility maximization genuinely underlies a wide range of phenomena. (Mäki, 2001b: p. 503). While derivational unification seems rather peculiar to an antirealist perspective about science, ontological unification is specifically realist.

2.2. The ontology of unrealistic assumptions

The problem of the ontology of unrealistic assumptions, implied in the subsections above, was also directly addressed. Musgrave (1981) has made a major criticism against Friedman's position, arguing that although some simplifications are acceptable, Friedman's use of a universal sentence (implying the homogeneity of assumptions), is however illegitimate. Based on this observation, Musgrave suggested a distinction between negligibility assumptions (i.e. assumptions that some factor has no effect on the phenomenon under study, or at least no detectable effect), and domain assumptions (i.e. negligibility assumptions that are subsequently refuted and become assumptions about the applicability domain of a theory) and heuristic assumptions (i.e. negligibility assumptions used in a first stage of modeling, which are later discarded in order to determine the results by comparison). Musgrave argues that negligibility assumptions and domain assumptions should be realistic because (in the first case) neglecting a non-negligible factor will affect the quality of predictions, and (in the latter case) a false domain assumption will impede on testing the theory. Mäki (2000) builds on Musgrave's typology, strengthening its main conclusions. He argues that Musgrave's wording of negligibility assumptions is ambiguous and it rather allows a discussion about the detectability assumptions (i.e. an assumption that an F factor has no detectable effect) than one about negligibility. Further Mäki relabeled domain assumptions as 'applicability assumptions' and heuristic assumptions as “first step assumptions”, and then, starting from Machlup (1955), he presented the distinction between core assumptions⁽⁸⁾ (fundamental postulates e.g. utility maximization) and peripheral assumptions (assumptions of specific conditions e.g. Galilei's vacuum assumption). Mäki argues that although Musgrave does not directly address the issue of these types of assumptions, the examples offered are rather about peripheral assumptions. Musgrave's typology can also be applied to core assumptions (Mäki, 2000: p. 329). To these clarifications, Hindriks (2005, 2006) adds the distinction between first-order assumptions (i.e. setting, for example, the absence or lack of effect of a factor F) and second-order assumptions (i.e. explaining why the first order assumption was

imposed). Additionally, Hindriks argues for introducing tractability or manipulability assumptions: “a second-order assumption according to which a first-order assumption is imposed in order to make a particular problem tractable. It is argued that a realist will want to relax a first-order assumption imposed for reasons of tractability as such assumptions are not even approximately true”. (Hindriks, 2006: p. 401) e.g. Newton’s one planet assumption.⁽⁹⁾ These assumptions are useful fictions (Hindriks, 2006: p. 414) and as argued by Kuorikoski, Lehtinen and Marchionni (2010) they are rather peripheral assumptions (Kuorikoski, Lehtinen, Marchionni, 2010: p. 547) serving as means of mathematical implementation of fundamental assumptions (i.e. factors that are in the model’s central mechanism). Mäki (2009a) argues however that there is a tension between tractability conventions and ontological convictions. When “pressures of tractability override important ontological considerations, and the values of formal rigor take over in shaping the focus and strategies of research” (Mäki, 2009a: p. 82) an important problem arises – the impossibility of accommodating modeling practice with scientific realism.

3. The near-neoclassical model of rational irrationality

The introductory section took note that the purpose of this article is to analyze the assumptions used by the rational irrationality model. In what follows I present this model along with its genesis. Then, using the framework presented in the previous section, I analyze the ontological status of the rational irrationality model’s assumptions, discussing its compatibility with a realist scientific program.

3.1. Preferences on beliefs

Akerlof and Dickens (1989) presented a model that incorporated cognitive dissonance in a neoclassical model. Simply defined, cognitive dissonance designates the process of aligning peripheral beliefs with fundamental ones: “In practice most cognitive dissonance reactions stem from peoples’ view of themselves as ‘smart, nice people’. Information that conflicts with this image tends to be ignored, rejected, or accommodated by changes in other beliefs.” (Akerlof, Dickens, 1989: p. 308). In order to use this idea into an economic model (i.e. with the utility maximization assumption) Akerlof and Dickens (1989) operationalized it into three sentences: Individuals not only have preferences over states of the world, but also over their beliefs about the states of the world; they have some degree of control over their beliefs and they are able, to a certain point, to manipulate their beliefs by selecting sources of information likely to confirm desired beliefs; once chosen, beliefs persists over time (Akerlof, Dickens, 1989: p. 307)⁽¹⁰⁾. Starting from Akerlof and Dickens⁽¹¹⁾ idea of modeling some degree of rationality with neoclassical tools, Caplan (2000, 2001a, 2004, 2006) proposed the

rational irrationality model as the main means of explaining the electoral preferences of citizens. The model is based, as in the case of Akerlof-Dickens' cognitive dissonance, on the idea that individuals have preferences over their beliefs. Caplan argues that rationality is not just a tool to achieve certain purposes but becomes an object of consumption that individuals (near-neoclassical agents) acquire according to its price: individuals "are trading wealth for irrationality" (Caplan, 2000: p. 91). Here, welfare must be understood in a broad sense, including not only consumption and portfolio value but also human capital, health or leisure: "in fact, one could just partition all arguments in the utility function into 'beliefs' and 'everything else', and use wealth as a synonym for the latter" (Caplan, 2000: p.193). This exchange between material wealth and mental well-being takes place under two important assumptions. The first refers to the fact that individuals perceive the cost of irrationality without bias (Caplan, 2000: p. 191) – which is actually the assumption that separates purely irrational individuals from the rational-irrational. The second concerns the concept of bliss belief – an irrational belief which will be consumed until satiation when its purchase cost is zero. (Caplan, 2001a: p. 314). Depending on this bliss belief and the cost (in terms of loss of material wealth) of deviations from rationality, individuals will consume more or less of irrationality. In Caplan's terms, "It is dangerous to think that poisonous substances are candy. It is dangerous to reject the theory of gravity at the top of the stairs. It is dangerous to hold that sticking forks in electrical sockets is harmless fun" (Caplan, 2006: p.119); on the other hand, it is not dangerous to have certain religious beliefs or to vote according to some economic or political sophisms. (Caplan, 2004: p. 471). On this last issue Caplan argues that in certain circumstances – most of the political circumstances – the price of irrationality is low: "The institutional structure of politics tends to peg the price of irrationality at zero." (Caplan, 2001a: p. 314) because "In politics, there is no private benefit of learning from experience, so irrationality can persist over time" (Caplan, 2001a: p. 327). The reason that the private cost of irrationality is "effectively zero" (Caplan, 2001b: p. 20) or at least very small is a classic one in public choice theory: the probability of individual decisiveness in mass elections is insignificant. Since individuals do not have biases about their preferences over beliefs, it is expected that they are aware of the lack of personal impact on election results. The aggregate impact of these private irrationalities can be though quite high: "It is cheap for an individual to irrationally underestimate the costs of war, but if enough people buy this opinion the result could be a disaster." (Caplan, 2001b: p. 20). Based on such observation, Caplan deduced political failure. Because voters have systematic biases (e.g. anti-market bias, anti-foreign bias, make-work bias and pessimistic bias)⁽¹²⁾ and since voting is an act without individual consequences, democracy will promote ineffective public policy⁽¹³⁾.

3.2. The ontological status of rational irrationality

As mentioned in the previous section, the rational irrationality model is a near-neoclassical one. (Caplan, 2000): individuals are the fundamental units of analysis (methodological individualism), they maximize utility and they are selfish at least in what concerns the selection of beliefs. The assumptions underlying the model have the main feature of neoclassical methodology – they are unrealistic. Therefore the discussion about the ontology of this model, about scientific realism and the acceptable unrealisticness of these assumptions is legitimate. Based on these observations, the arguments mentioned in the paper's abstract and in the introductory section are formulated in the lines that follow.

3.2a. The mixed-up ontology of rational irrationality

Caplan's model provides conflicting information about the descriptiveness of the central assumption of the rational irrationality model – the unbiased perception of the costs of irrationality. On the one hand, Caplan presents numerous instances in which he seems to be concerned about the psychological content of his model and about the degree to which rational irrationality is descriptively valid (i.e. not only behaviorally but also psychologically). For example, Caplan (2006) notes that other authors did not follow a path similar to that of rational irrationality because of „its psychological implausibility” (Caplan, 2006: p. 125). He then argues that rational irrationality is a psychologically credible notion. Also, in cases where rational irrationality is behaviorally indistinguishable from other concepts⁽¹⁴⁾, Caplan argues that the difference must be however, made. For example, in the case of the difference between rational irrationality and rational ignorance, Caplan argues that “while rationally ignorant individuals admit they are ignorant, rationally irrational individuals believe that they know the truth” (Caplan: 2001a: p. 315). A similar argument is given for the distinction between rational irrational voters and expressive voters: “The parallels with rational irrationality are clear. [...] The key difference is the mechanism. In expressive voting theory voters know that feel-good policies are ineffective. [...] In contrast, rationally irrational voters believe that feel-good policies work” (Caplan, 2006: pp. 138-139). Furthermore, Caplan discusses an example of Brennan and Lomasky (1985). Expressive reasons, argue the two, can lead to disastrous results such as the declaration of war: the individual indecisiveness in mass elections cancels the cost of expression of national pride which, most likely, individuals would not express if their vote would be decisive. Caplan shows that while Brennan and Lomasky' 'story' is logically possible, it 'comes off as odd' unless we relax the rationality assumption: “How many vocal hawks would admit to themselves that war leads to devastation?” (Caplan, 2006: p. 139). All these distinctions seem to be less about agents' behavior⁽¹⁵⁾ and more about their psychological features. In the first case

(that of difference from rational ignorance) the differences are not just psychological⁽¹⁶⁾. In the second case the behaviors are the same, and the difference seems to lie only in the psychological characteristics of actors⁽¹⁷⁾. This concern for psychological descriptiveness of rational irrationality results also from the way Caplan writes about the importance of introspection⁽¹⁸⁾ – something very unlikely to have instrumentalist connotations. The concern for psychological descriptiveness is also visible in the description of rational irrationality as having tacit cognitive steps (2006: p. 126). Also, the mention of the cognitive process behind these steps reinforces the idea of Caplan's concern for the psychology of rational irrationality.

On the other hand, despite all the above examples, there are several instances where rational irrationality is actually presented as being rather behavioral than psychological. For instance, Caplan (2000) notes that “As with utility theory in general, one should not read too much psychological content into this choice model” (Caplan, 2000: p. 194). This idea is strengthened in (2001b): we should give no “specific psychological, anthropological, or sociological interpretation to choice over beliefs” (Caplan, 2001b: p. 10) and then in (2006): “The rational irrationality model is consistent with a much broader range of observations than standard neoclassical model” (Caplan, 2000: p.199). The lack of psychological content, and also the concept of compatibility with factual observations can be viewed as instrumentalist commitments. Therefore, Caplan seems to follow Friedman's tradition of relying on behavioral assumptions consistent with observable facts: individuals would behave *as if* they would be rational-irrational. The lack of conflict between observation and theory seems to be a condition of a rather instrumentalist origin. Moreover, the concern for the psychological descriptiveness of rational irrationality seems to conflict with Caplan's choice to introduce it into a neoclassical methodological framework. Complete, continuous, transitive preferences, or the idea of utility maximization are not psychologically descriptive⁽¹⁹⁾ and they seem to have a rather mathematical tractability function, which Caplan points at in several instances. A first example is the use of well-defined preferences: “Analyzing behavior without well-ordered preferences is quite difficult; but given well-ordered preferences, analyzing irrational beliefs is - it will be argued - a manageable task” (Caplan, 2000: p. 193). The reason behind their use appears to be linked to their tractability – a virtue for building deductive models. The distinction between cognitive and motivational bias seems to lead to the same idea: “The motivational depend on the emotions, but the cognitive do not. But the rational irrationality model treats them symmetrically for analytical purposes.” (Caplan, 2000: p. 194) Therefore, biases are artificially homogenized by reasons of modeling constraints and not by psychological fidelity reasons. Similar, but even more important, Caplan points out that giving up the rational

expectations methodological framework “provides only a semantic victory” because such a waiver has “a very high cost. So once you lower the threshold of rationality, you can no longer safely build on the standard rational actor theorems. You have to go back to square one to save a word” (Caplan, 2006: pp. 99-100). The reasons for modeling preferences over beliefs in this way are therefore not connected with an attempt to explain behavior as best it can be explained but rather with the attempt to explain it with some particular tools.

The picture that can be drawn from the ways in which Caplan presents the rational irrationality is mixed. On one hand, there is a significant concern for the psychological descriptiveness of the model’s assumptions. On the other hand, it is asserted that this is irrelevant. At the same time, Caplan offers reasons for modeling rational irrationality in the neoclassical methodological framework but then they are quickly opposed by statements such as the following: “Unlike ignorance, irrationality allows a wide range of outcomes. Many see the absence of a unique prediction as a defect, or sign of intellectual sloth. I do not. As Richard Thaler pointedly asks, ‘Would you rather be elegant and precisely wrong or messy and vaguely right?’” (Caplan, 2006: p. 113). This quote seems to put the consistency of Caplan’s claims about rational irrationality in a serious mess. If disorder is indeed justified, then Caplan’s arguments for adopting the neoclassical framework seem less explicable. If order and having unique predictions are not important, then why not go back “to square one”? (Caplan, 2006: p. 100), totally rejecting the psychologically undescriptive neoclassical framework? Why should we mix cognitive and motivational bias and why should we use well-defined preferences? To these questions it should be added the observation that rational irrationality seems to claim no more than the reactivity at the price of irrationality. It could be argued, therefore, that utility maximization is an unnecessary addition, and if Caplan’s concern would be purely realist, then rational irrationality could simply be reduced to the theory of origin – the cognitive dissonance – or could be accommodated with higher psychological content economic models⁽²⁰⁾.

3.2b. The functions and effects of rational irrationality unrealistic assumptions

In the previous section, I mentioned the tractability functions of some of the unrealistic assumptions used by Caplan. In section 3.2b the assumptions of the rational irrationality model will be carefully studied, examining the extent to which the simplifications made by Caplan are compatible with a realist scientific program. I begin by clarifying the logical structure of the model using the case of electoral preferences.

(a1): the individual decisiveness probability (p) in mass elections is very low (the model’s only structural assumption);

(a2a): individuals maximize utility of a mix made by mental and material welfare (their preferences are complete, reflexive, transitive, continuous, monotonous, and convex⁽²¹⁾);

(a2b): individuals accurately estimate the cost of irrationality in terms of loss of material welfare (implied by a2a);

(a2c): individuals accurately estimate the probability of being decisive in mass elections (implied by a2b);

(a2d): individuals are selfish in selecting their beliefs;

(a2e): individuals are altruists when they vote.

The first statement that can be made is that the Musgrave-Mäki-Hindriks typology of assumptions focuses on assumptions' functions, but they do not discuss the possibility that an assumption could have several simultaneous roles (Ungureanu, 2013). Starting from Hindriks' (2005, 2006) distinction between first order and second order assumptions it could be argued that assumptions are rarely accompanied by meta-statements the type of 'this assumption has the function of ...' For this reason we can only presume which functions an assumption fulfills in a given model and therefore it could be argued that what we can really observe are the effects of that assumption.⁽²²⁾ (Ungureanu, 2013) Once we accept that in the current modeling practice such meta-statements which specify the functions that assumptions have are rare and *once referring to their effects seems reasonable*, we can see that an assumption could have multiple simultaneous effects. For example, (a2d) can be read as 'in selecting beliefs, any reason other than selfishness can be neglected' – a negligibility effect. At the same time, however, this assumption has a tractability effect. It produces a uniform model world (i.e. actors do not have multiple motivations) allowing clear, unique predictions. (a2e has the same effects) Another observation that can be drawn is that the tractability effects cannot be limited to peripheral assumptions (Ungureanu, 2013). Core assumptions may have tractability effects as well. The best example is the very assumption of utility maximization. This is fundamental to neoclassical methodology for reasons not concerning ontological fidelity but mathematical tractability. Caplan even argues that to abandon this principle would make the analysis very difficult (Caplan, 2000: p. 193) and would send the researcher back to square one (Caplan, 2006: p. 100). So the reason Caplan has in mind when modeling rational irrationality in the utility maximization framework, is to enhance tractability effects. Admitting, however, that there may be simultaneous effects, we can rewrite the utility maximization assumption as a negligibility assumption: 'we can safely neglect actors who use other rational choice criteria (e.g. minimax regret) and actors that are purely irrational'. Starting from this interpretation of utility maximization but also from the mathematical tractability effect that this assumption has (including the uniformity effect), it has to be

recalled the tension between tractability and ontological considerations (Mäki, 2009). Here, this is not only an intra-model but also an intra-assumption tension. The question is of course that of deciding whether tractability considerations should override the ontological considerations in this case. If in the case of alternative rational choice criteria there is some empirical support for their negligibility⁽²³⁾, in the case of irrationality this is not as clear. Suppose, however, that tractability reasons are not interfering in this case with the causal mechanism of the target world, and we could safely turn our attention to another assumption of the model i.e. (a2c). This can also be rewritten as a negligibility assumption: ‘voters who do not know the value of p can be neglected’ and it also has tractability effects – it assumes uniformity in the model world. The problem with this unrealistic assumption is that it has questionable empirical support and, in this case, ontological considerations seem to be subordinated to those of tractability. (a2c) is also related to (a2b), which is related to (a2a). If individuals are not able to understand the concept of probability, or do not know the probability of an event (a2c) then they cannot accurately estimate the cost of irrationality (a2b) and in effect they cannot maximize expected utility (a2a) leading to the model’s failure to produce predictions. Beside the structural assumption (a1) which is unproblematic⁽²⁴⁾, all of the other assumptions are unrealistic and include tractability reasons. For instance, they all have the uniformity effect mentioned above. Of course, tractability assumptions are not problematic in themselves. They become so only when they are not only “strategic falsehoods” (Mäki, 2009: p. 78). In the case of a2a-a2e there are reasons to believe that tractability considerations interfere with ontological considerations making Caplan’s model suspect of having an instrumentalist approach.

4. Discussion and conclusion

Caplan's model seems to be torn apart between two worlds. On the one hand he insists on the psychological content of rational irrationality; on the other hand it reflects opposing motivations regarding modeling constraints in a particular methodological framework. At times Caplan seems more concerned about being ‘vaguely’ right than, being ‘precisely’ wrong (Caplan, 2006: p. 113). In others, however, the interest in the issues of tractability is more visible⁽²⁵⁾. This dissociated identity of the rational irrationality model makes it difficult to fit in a realist or anti-realist tradition. In addition, Caplan's view on the neoclassical-behavioral debate “as a disagreement within ‘normal science’ about parameter value.” (Caplan, 2000: p. 191) increases the confusion. The methodological foundations of the two approaches have little in common and the methodological distinction between fixing a type of rationality (neoclassical economics) and

discovering how individuals actually behave (behavioral economics) are fundamentally different strategies for generating knowledge. If in the first case, beginning with Friedman, economists have legitimized themselves from anti-realistic justifications, in the second, the ontological status of the assumptions used by neoclassical economics was the point of divergence. Additionally, Caplan fails to provide ontological grounds for using the neoclassical framework in explaining electoral behavior and not alternative methodological frameworks.

Notes

- (1) The reason for receiving this label was Popper's (1962) definition of instrumentalism as an attitude that theories are meant only to make predictions. Mäki (2009b) and Blaug (2009) did not though join the quasi-consensus in classifying Friedman as instrumentalist even though Friedman himself defined his position as such (Blaug, 2009: pp. 350-351).
- (2) Rational irrationality could be applied to other issues as well but Caplan heavily insisted on electoral preference formation.
- (3) Mäki (1989) argues that the term 'realism' is legitimately used when it designates a variety of philosophical theses. When we wish to refer to different properties of linguistic representations or economic theories and their parts it is preferable to use the term 'realisticness'. To avoid confusion between the two meanings of the term 'realism' I use Mäki's choice of words 'unrealisticness of assumptions'.
- (4) Likewise, the leaves of a tree move *as if* they maximize the light they receive, and pool players behave *as if* they knew the mathematics necessary for identifying the optimal paths and angles to hit the ball (Friedman, 1953: pp. 20-21).
- (5) The idea is first expressed by Putnam (1975).
- (6) i.e. electrons, quarks.
- (7) Mäki offers Galilei's law of free falling bodies which is based on the absence of air pressure or other forces. These assumptions are clearly unrealistic, but they are used to isolate the impact of gravity alone on the falling bodies. (Mäki, 2009a: p.79) They have a tractability function serving as "formal auxiliaries rather than as distortions of an actually held deeper worldview." (Mäki, 2009a: p. 82).
- (8) Another way is to determine the causal core of a model by using robustness analysis as developed by Ungureanu (forthcoming).
- (9) Hindriks explains that Newton lacked the necessary mathematical apparatus needed and therefore, for tractability reasons, he assumed the existence of just one planet.
- (10) For example, Akerlof and Dickens noted that in experiments (Gass, 1964), it was observed that students who accepted to apply electric shocks to some experimental victims, changed their views on these victims – they chose to resolve the conflict between self-image and their cruel behavior by having a bad opinion about the victims – a view that they previously did not have (changing of beliefs). Since the conflict between cruel behavior and self-image is relatively independent of time, it is expected that this will be a long-lasting effect (preferences stability over time).
- (11) The details of Caplan's model are irrelevant for my purposes here. It is only worth mentioning that the premises presented above were used to explain the necessity of adopting labor safety regulations.

- ⁽¹²⁾ Caplan (2006) identifies four systematic voter biases: anti-market bias (i.e. voters tend to believe in conspiracy rather than in the role of supply and demand forces when it comes to fixing the market price), the bias against foreigners (i.e. individuals tend to favor protectionist policies even if serious evidence that they are ineffective is available), the make-work bias (i.e. people tend to believe that employment is good for economy in itself) and the pessimistic bias (i.e. individuals systematically perceive that things in the economy are bad and will not get any better).
- ⁽¹³⁾ The details of Caplan's conclusions and extensive empirical research will be omitted here.
- ⁽¹⁴⁾ e.g. X behaves *as if* he is rational irrational but also *as if* he is an expressive voter.
- ⁽¹⁵⁾ Friedman's tradition is concerned with behaviors not with the psychological fidelity of assumptions.
- ⁽¹⁶⁾ Caplan shows that rational ignorant individuals have unbiased behaviors - their errors are not systematic - while the errors of rational-irrational individuals are systematically biased - the two concepts are both behaviorally and psychologically distinct.
- ⁽¹⁷⁾ In fact the two concepts are analytically indistinguishable, at least in terms of voting behavior.
- ⁽¹⁸⁾ "One thing my introspection tells me is that some beliefs are more emotionally appealing than their opposites. For example, I like to believe that I am right. It is worse to admit error, or lose money because of error, but error is disturbing all by itself [...] Introspection is a fine way to learn about your own preferences" (Caplan, 2006: p. 117).
- ⁽¹⁹⁾ And as shown by numerous experiments in behavioral economics (Tversky and Kahneman 1981, Kahneman, Slovic and Tversky, 1982), they are neither behaviorally descriptive.
- ⁽²⁰⁾ An alternative closer to economists' concerns could be the choice to model preferences over beliefs in Simon's (1995) bounded rationality framework or even in the behavioral economics framework. None of these options are though explored in this paper.
- ⁽²¹⁾ Caplan explicitly assumes only the existence of well-defined preferences (Caplan, 2001b, p. 7), but using arguments illustrated by indifference maps (Caplan, 2000: p. 193, 2001 p. 313) and evoking the methodological framework of "rational expectations" (Caplan, 2001a: p. 313) involves all these conditions imposed on preferences over beliefs.
- ⁽²²⁾ In other words, to discuss about functions (roles intended by the researcher) is not operable.
- ⁽²³⁾ The research conducted by Blais, Young, Fleury, Lapp (1995) shows that minimax regret "is a little more than a rationalization on the part of those having a strong sense of duty to vote" (Blais, Young, Fleury, Lapp, 1995: p. 827). It is also informative the anecdotal description that Dhillon and Peralta (2002) offer for this choice criteria: "a minimax agent should never cross the street." (Dhillon, Peralta, 2002: p. F338).
- ⁽²⁴⁾ Most democracies have a large enough demos, so the value of p is indeed very low.
- ⁽²⁵⁾ Indeed it can be even seen as a concern for explanatory unification "the application of the economic way of thinking to irrationality can also be seen as a striking vindication of economic imperialism." (Caplan: 2001b: p. 22).

References

- Aronson, E. (1979). *The Social Animal*, W.H. Freeman, San Francisco
- Bear, D., Orr, D. (1967). "Logic and expediency in economic theorizing", *Journal of Political Economy*, 75, pp. 188-96
- Becker, G. (1976b). The Economic Approach to Human Behavior, in: Gary Becker, *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press, pp. 3-14

- Blais, A., Young, R., Fleury, C., Lapp, M. (1995). "Do People Vote on the Basis of Minimax Regret?", *Political Research Quarterly*, Vol. 48, No. 4, pp. 827-836
- Blaug, M. (2009). "The debate over F53 after fifty years", in Maki (Ed.), *The Methodology of Positive Economics Reflections on the Milton Friedman legacy*, Cambridge University Press (2009b)
- Boland, L.A. (1979). "A critique of Friedman's critics", *Journal of Economic Literature*, 17, pp. 503-22
- Brennan, G. & Lomasky, L.E. (1985), "The Impartial Spectator Goes to Washington: Toward a Smithian Theory of Electoral Behavior". *Economics and Philosophy*, 1, pp. 189-211
- Caldwell, B. (1980). "A critique of Friedman's methodological instrumentalism", *Southern Economic Journal*, 47, pp. 366-74
- Camerer, C. (1987). "Do Biases in Probability Judgement Matter in Markets? Experimental Evidence", *American Economic Review*, 77(5), pp. 981-97
- Caplan, B. (2000). "Rational Irrationality: A Framework for the Neoclassical-Behavioral Debate". *Eastern Economic Journal*, 26(2), pp. 191-211
- Caplan, B. (2001a). "Rational irrationality and the microfoundations of political failure", *Public Choice*, 107, pp. 311-331
- Caplan, B. (2001b). "Rational Ignorance versus Rational Irrationality". *Kyklos*, 54(1), pp. 3-26
- Caplan, B. (2001c). "Libertarianism against Economism: How Economists Misunderstand Voters, and why Libertarians should Care", *The Independent Review*, Vol. 5, No. 4, pp. 539-563
- Caplan, B. (2001d). "Probability, Common Sense, and Realism: A Reply to Hülsmann and Block", *The Quarterly Journal of Austrian Economics*, Vol. 4, No. 2: 69-86
- Caplan, B. (2003a). "The Logic of Collective Belief", *Rationality and Society*, Vol. 15(2), pp. 218-242
- Caplan, B. (2003b). "Stigler–Becker versus Myers–Briggs: why preference-based explanations are scientifically meaningful and empirically important", *Journal of Economic Behavior & Organization*, Vol. 50, 391-405
- Caplan, B. (2006). "The Myth of the Rational Voter: Why Democracies Choose Bad Policies", Princeton University Press
- Dhillon, A., Peralta, S. (2002). "Economic Theories of Voter Turnout", *The Economic Journal*, Vol. 112, No. 480, pp. F332-F352
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford
- Friedman, M. (1953). "The Methodology of Positive Economics", in Maki (ed.), *The Methodology of Positive Economics Reflections on the Milton Friedman legacy*, Cambridge University Press (2009b)
- Hands, W. (2003). "Did Milton Friedman's methodology license the Formalist Revolution?", *Journal of Economic Methodology* 10:4, pp. 507-520
- Hindriks, F. (2005). "Unobservability, Tractability and the Battle of Assumptions", *Journal of Economic Methodology* 12:3, pp. 383-406
- Hindriks, F. (2006). "Tractability assumptions and the Musgrave-Mäki Typology", *Journal of Economic Methodology*, 13:4, pp. 401-423
- Kahneman, D., Slovic, P., Tversky, A. (1982). *Judgement under Uncertainty: Heuristics and Biases*, Cambridge University Press
- Kuorikoski, J., Lehtinen, A., Marchionni, C. (2010). "Economic Modelling as Robustness Analysis", *British Journal for the Philosophy of Science*, 61, pp. 541-567
- Machlup, F. (1955). "The Problem of Verification in Economics", *Southern Economic Journal*, Vol. 22, No.1, pp. 1-21

- Mäki, U. (1986). „Rhetoric at the expense of coherence: a reinterpretation of Milton Friedman’s methodology”. *Research in the History of Economic Thought and Methodology*. 4, pp. 129-43, Greenwich, CT: JAI Press
- Mäki, U. (1989). “On the problem of realism in economics”, *Ricerche Economiche*, 43, pp. 176-198
- Mäki, U. (1994). “Reorienting the assumptions issue”, in Backhouse, R. (ed), *New Directions in Economic Methodology*, London: Routledge, pp. 236-256
- Mäki, U. (2000). “Kinds of Assumptions and Their Truth: Shaking an Untwisted F-Twist”, *Kyklos*, Vol. 53, Fasc. 3, pp. 317-336, 317
- Mäki, U. (2001). “Explanatory Unification: Double and Doubtful”, *Philosophy of the Social Sciences*, 31:488-506
- Mäki, U. (2005). “Reglobalizing Realism by Going Local, or (how) should our Formulations of Scientific Realism be Informed about the Sciences?”, *Erkenntnis*, 63, pp. 231-251
- Mäki, U. (2009a). “Realistic Realism about Unrealistic Models”, in H. Kincaid, D. Ross (eds.), *The Oxford Handbook of Philosophy of Economics*, Oxford University Press
- Mäki, U. (2009b). “Reading the methodological essay in twentieth-century economics: map of multiple perspectives”, in Mäki (Ed.), *The Methodology of Positive Economics Reflections on the Milton Friedman legacy*, Cambridge University Press
- Mullainathan, S., Washington, E. (2009). “Sticking with Your Vote: Cognitive Dissonance and Political Attitudes”, *American Economic Journal: Applied Economics* 2009, 1:1, pp. 86-111
- Musgrave, A. (1981). “Unreal Assumptions in Economic Theory: The F-Twist Untwisted”, *Kyklos*, Vol. 34, Fasc. 3, pp. 377-387
- Popper, K. (1962). “Three Views concerning Human Knowledge”, in Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Chapter 3, Basic Books Publishers, pp. 97-119
- Psillos S. (2005). *Scientific Realism: How Science Tracks Truth*, Routledge: London and New York
- Putnam, H. (1975). *Mind, Language and Reality*, Cambridge University Press
- Simon, H.A. (1995). “Rationality in Political Behavior”, *Political Psychology*, Vol.16, No.1, Special Issue: *Political Economy and Political Psychology*, pp. 45-61
- Tversky, A., Kahneman, D., (1981). “The framing of decisions and the rationality of choice”, *Science*, 211, pp. 453-458
- Ungureanu, M. (2013). “Uniformity Assumptions and the Musgrave-Mäki-Hindriks Debate”, *Innovative Issues and Approaches in Social Sciences*, Vol. 6, No. 2
- Ungureanu, M. (forthcoming). “Robustness of Rationalistic Models of Bureaucracy”