António Zilhão

COGNITION AND RATIONALITY: WRITING STRAIGHT WITH CROOKED LINES?

Abstract

Four different approaches dominate the modern discussion around the topic of how best to define human rationality. These are the following: 1) Unbounded Rationality (UR); 2) Optimization under Constraints (OuC); 3) Heuristics and Biases (H&B); 4) Ecological Rationality (ER). Typically, proponents of approaches 3) and 4) criticize the models put forth by the proponents of approaches 1) and 2) for their cognitive unrealism. However, many ethologists contend that it makes sense to account for data gathered in animal behavior research along the lines of precisely these models. Elaborating upon this claim, Stanovich suggested that models of the kinds 1) and 2) are more appropriate to account for the behavior of creatures endowed with simple cognitive architectures rather than to account for the behavior of humans. Moreover, according to Stanovich's own new approach, it is the cognitive complexity of humans rather than their computational limitations that makes them (partially) irrational. Could he be right? Following a suggestion made by Kacelnik (2006), I will contend that, in order to try to find out an answer to this question, one needs to realize that the term "rationality" in use in this debate has to be understood as referring to, at least, three distinct properties; and that, once one analyzes each of them properly, most of the above-mentioned approaches, including Stanovich's, reveal themselves to be untenable.

1. Models of Human Rationality

Since the beginning of the second half of the twentieth century, the philosophical, psychological and social-scientific debate concerning how best to define human rationality has been revolving around the following four main views: 1) Unbounded rationality (UR); 2) Optimization under constraints (OuC); 3) Heuristics and biases (H&B); and 4) Ecological rationality (ER).

Let me review them briefly.

1.1 *Unbounded Rationality (UR)*

UR is considered to be the classical view in rationality theory. Famous holders of this view are, e.g., von Neumann and Morgenstern (1944), Arrow (1951), Friedman (1953), Savage (1954) or Jeffrey (1983). According to their standpoint, normal (i.e., rational) human action is to be viewed as some sort of empirical instantiation of a model of normative optimization (rational choice theory). This approach views agents as the holders of sets of preferences and assumes these sets to be well behaved in terms of a certain number of axioms of choice (e.g., completeness, transitivity, independence of irrelevant alternatives, sure-thing principle). It then proves that non-

self-defeating agents, thus characterized, behave in order to maximize individual expected utility.

Typically, UR models do not take into account any constraints (e.g., time constraints or cognitive processing limitations). This is, of course, the reason why the term 'unbounded' is used in their characterization.

This disregard from time or procedural constraints entails two unwelcome consequences though. First, in many situations, in the absence of such constraints, the problem space remains unlimited and, therefore, a UR model cannot even be specified (cf. Chase et al. 1998, p. 207). Second, algorithms implementing those models that can be specified tend to generate unavoidable problems of computational complexity and intractability.

The moral usually drawn from these consequences is twofold: first, UR models ought to be seen, first and foremost, as regulatory ideals; second, empirical cases of human decision-making and action ought to be considered to be only approximate instantiations to these ideals.

1.2.*Optimization under Constraints (OuC)*

OuC is a view trying to combine the classical approach to rationality theory with a concern with, at least, some of the bounds within which decisions in the real world have to be made. Thus, OuC theoreticians try to capture the structure of normal (i.e., rational) human action through the introduction of a number of constraints into the design of traditional models of normative optimization.

These theoreticians admit being divided into two groups, according to the nature of the constraints they choose to introduce in their models.

To the first group belong those claiming that normal human action is best viewed as an instantiation of a model of optimization that is subject to a number of time, cost, or information processing constraints (see, e.g. Stigler 1961). The holders of this view have developed optimization algorithms that are supposed to represent the way the *mind* takes into account variables such as, e.g., the cost of information gathering. According to such algorithms, search should terminate when its costs outweigh its benefits.

To the second group belong those claiming that normal human action is best viewed as an optimal adaptation of the agent's behavior, given its goals, to certain relevant aspects of the decision environment. Their idea is to use optimization algorithms in order to predict human behavior from the structure of the *environment* rather than from the structure of the mind. This is the view usually called "Rational Analysis" (cf. Anderson 1990, 1991).

However, no matter the group they belong to, the algorithms put forth by OuC theoreticians tend to be computationally extremely complex. That is, they tend to suffer from the very same intractability problems – typical of UR models – that they were intended to overcome in the first place. As a matter of fact, J. R. Anderson himself could not help confessing that (Anderson 1991, p. 474):

> The major complication in this step is analytic tractability.

1.3. *Heuristics & Biases (H&B)*

Proponents of the H&B approach start from the claim, originally made by Herbert Simon (1969), that the human mind is inescapably bounded; as a consequence, they suggest that normal human action is determined mainly by non-rational heuristics and biases the human cognitive apparatus is stuck with in the course of the evolutionary process. These heuristics and biases, frequently called "quick and dirty", are to be seen as simple cognitive shortcuts that direct us in our ways of acting; moreover, they are supposed to be independent from experience and very hard to overcome by learning. One way of conceiving of them is as cognitive analogs of the visual illusions studied in the psychology of perception (such as the famous Müller-Lyer illusion). Holders of this view are, *inter alia*, Tversky (1969, 1974, 1975, 1981), Kahneman (1982, 2011), Piatelli-Palmarini (1994) or Lichtenstein and Slovic (2006).

Thus, according to these researchers, the "ways of man" contradict, to a large extent, the prescriptions of classical rational choice theory. Man should therefore be considered to be a non-rational creature (partially, at least). The job of rationality theorists, in turn, should be to strive to identify these heuristics and biases and to suggest methods for helping people becoming aware of their own irrationality and correct it.

This approach has in the meantime evolved into a dualistic model of reasoning and action, according to which the human cognitive apparatus should be seen as being endowed with two systems of reasoning rather than one. One of these systems (System 1) is supposed to be fast and to include a large stock of non-rational quick and dirty heuristics; the other system (System 2) is supposed to comprise an important set of normatively rational inferential procedures, to be slow and to exist in close association with language and consciousness (cp. e.g. Sloman 1996, Kahneman 2011). People are supposed to act in agreement with System 1 when they are not acting reflectively; under circumstances in which careful reflection is possible, System 2 steers action. There are many cases, however, in which both systems get mobilized and compete for dominance.

1.4. *Ecological Rationality (ER)*

Also elaborating upon H. Simon's original approach, ER is the view according to which rationality is to be equated with adaptability.

Similarly to H&B theoreticians, ER supporters accept the idea that the human mind is inescapably bounded. Thus, they agree with H&B theoreticians that normal human action is frequently the outcome of the triggering of non-normative heuristics. But they contend that the H&B research program has failed to find and properly identify any of them.

Rather than "quick and dirty", ER theoreticians prefer to call the heuristics they claim the human mind is genuinely equipped with "fast and frugal". This replacement of the negative term 'dirty' with the more positive term 'frugal' is, obviously, not merely terminological. As a matter of fact, ER researchers accuse H&B theorists of having neglected one of the two key aspects of H. Simon's original approach. This is the aspect that human rationality, besides being bounded, is also 'situated'. What Simon meant by the use of this term was the idea that human rational behavior needs to be understood by reference to the environment within which it is supposed to be deployed. ER theorists think this aspect of Simon's original thinking needs to be strongly emphasized.

Thus, dissimilarly to H&B theoreticians, ER proponents state that, within its proper environment, and under the right sort of circumstances, the display of any of the heuristics they study admits being seen as a more robust way of dealing with an uncertain and complicated world than highly complex optimization tools. This means that, according to their standpoint, resorting to a fast and frugal heuristics might frequently be the rational (i.e., adaptive) way to go.

The main holders of this view are G. Gigerenzer, P. Todd and The ABC Group (1999).

2. The Debate

2.1. *Round 1*

Proponents of the H&B or ER views typically criticize UR or OuC models for being descriptively inadequate at the behavioral level and highly unrealistic at the cognitive level. More particularly, they put forth the objections (A) and (B) below.

Objection (A): The axioms (e.g., transitivity or independence) from which normative models of rational behavior have been derived have been shown by experimental economists and cognitive psychologists not to be true of concrete human subjects in multiple situations (cf., e.g., Allais 1953, Tversky 1969, 1975).

Objection (B): The complexity involved in the implementation of these models outside of simplified toy worlds is staggering; therefore, when matters are considered from an evolutionary perspective, the cognitive processing of such a complexity would be too costly and largely inefficient.

Objection (B) is, of course, just a rewording, wrapped up in evolutionary considerations, of the two unwelcome consequences of rational choice theory already diagnosed in section 1.1.

Supporters of the H&B or ER standpoints thus contend that expecting real people to routinely conform to such norms, even if only approximately, requires not only disregarding important results in Cognitive Psychology, but also assuming that their minds contain a kind of superdupercomputer similar to the one needed to equip the mind of a Laplacean demon. This is, of course, unacceptable. Thus, models based on such expectations and assumptions should be viewed as describing the would-be cognitive and behavioral capabilities of god-like creatures and not those of real world biological creatures, such as humans.

2.2. *Round 2*

In general, UR or OuC supporters respond to these criticisms along the same lines that were famously put forth by Friedman (1953) more than sixty years ago. These are the following:

Reply to Objection (A): The sort of criticism instantiated by this objection falls prey to a basic logical fallacy, namely, the fallacy of the negation of the antecedent. This reply has a Popperian flavor: it implies that theories are to be assessed by the success of their predictions only. From this perspective, empirically driven axiom testing is basically pointless (cf. also Davidson 1980, 1995).

Reply to Objection (B): The sort of criticism instantiated by this objection fails to see that the assumptions backing up the use of the axioms of rational choice as the

foundation for a theory of human action are actually biologically sound. These assumptions are the following. First, both cognition and behavior are adaptive; second, what adaptation amounts to is optimization. Criticisms based on unproven allegations regarding the details of cognitive implementation (about which nobody knows much, really) are therefore both misplaced as well as theoretically detrimental. These criticisms forget that evolution *writes straight with crooked lines*. Once again, J. R. Anderson put the point nicely (cf. Anderson 1991, p. 483):

> It is in the spirit of a rational analysis to prescribe what the behavior of a system should be rather than how to compute it.

2.3. *Round 3*

Obviously, the dialectic does not stop here. ER theoreticians, for instance, claim that what they are looking for is an *integration* of the question of *why* it is that humans behave in the way they do with the question of *how* it is that they manage to do what they do. Their ambition is therefore meant to be explanatory in a *deeper* sense, namely, in the sense in which an explanation of a natural phenomenon should provide us with some understanding of the underlying causal mechanisms producing it and not only with the ability for making successful predictions. Seeing things from this perspective, computational tractability and feasibility seem to be unavoidable boundaries for the cognitive implications (the *crooked lines*) of a behavioral theory. Another researcher supporting this line of inquiry is, e.g., Jonathan St. Evans, according to whom the only real explanatory question involving human cognition is the question of determining what are the mechanisms that our phenotypes have developed in order to reach cognition's goals (Evans 1991 p. 493).

Deeper explanations should also have better predictive value though (that is, they should *write straight* too). Thus, ER theorists devised various cognitive experiments aimed at showing that the outcome of the deployment under the appropriate circumstances of the fast and frugal heuristics they claim to have identified match human behavioral facts better than the predictions made in agreement with UR, OuC or H&B standpoints.

However, the right interpretation of these experiments is a matter of controversy; and, on the other hand, at least some OuC theorists contend that, in this context, the claim that alleged mechanistic explanations are, in any meaningful sense, deeper is purely arbitrary (cf. Anderson 1991, p. 512). It is not my purpose to lead this discussion into a general discussion around the nature of psychological explanation though; so let me close this section here.

3. Unexpected Findings and a Possible "Third Way"

More recently, Stanovich (2013) suggested a fifth view in rationality theory. I will call it "Brute Rationality" (BR) for reasons that will become apparent below. BR puts together elements of the earlier views and combines them in a new direction. This new direction was, in turn, suggested to Stanovich by a series of earlier remarks made by Searle (2001), on the one hand, and by the consideration of unexpected findings he came across when reviewing animal behavior literature, on the other hand.

3.1. *Ultimatum Game with Chimpanzees*

These unexpected findings are undoubtedly important. Thus, and leaving Searle aside, let me make a small detour through some of them before summarizing the essentials of Stanovich's BR approach to rationality. These findings consist of several experimental results obtained in animal behavior research that seem to suggest that, in a number of circumstances, non-human animals behave rationally in situations in which humans do not. That is, frequently, humans seem to be *less* rational than other animals (e.g., Kagel, C. J. 1987).

Jensen, Call and Tomasello (2007) provide us with one good example of such work. They were able to devise a version of the famous ultimatum game that was accessible to chimpanzees. They then compared the results of this version of the game presented to the chimpanzees with the results obtained in the human version of the same game by experimental economists.

The outcome of this comparison was the following: chimpanzees always behaved as rational maximizers whereas humans, in general, do not. As is well known, human responders tend to reject low offers because, in their own words, they are "unfair". In fact, most human proposers anticipate that responders will respond this way and, in general, end up making fair offers. Not so with chimpanzees. Proposers almost always made selfish offers and responders almost always accepted anything that was greater than nothing, in agreement with the prescriptions of rational choice theory.

Now, when read either from the perspective defined by UR or OuC standpoints or from the cognitively inspired criticisms of them put forth by the proponents of the H&B and ER views these findings sound paradoxical.

3.2. *"Economic Animal" vs. "Irrational Man"?*

As mentioned above, impressed by these results, Stanovich advanced his own view – BR. In a sense, BR may be regarded as a sort of 'third way' in the rationality theory debate. Although it collects elements from the UR and OuC camp, on the one hand, and from the H&B and ER camp, on the other hand, it remains distinct from both of them.

According to Stanovich's new view, it is not god-like creatures but brutes that admit being best described by traditional optimization models of rationality. That is, being indisputably adaptive, non-human animal cognition is the proper target for being the subject of a Friedman-like optimization approach. Humans, on the other hand, fail to be rational; however, Stanovich claims that humans fail in this endeavor not because their cognitive make-up lacks the computational resources needed to implement highly complex optimization procedures but, rather, because their *own* cognitive complexity makes it harder for them to conform to the strictures of rational choice.

Therefore, so the story goes, when *unhindered* by human mental complexity, natural selection shapes the decision algorithms contained in the simple brains of animals into those whose output is the behavior typical of "economic man". Humans, in turn, seem to have been subject to strange evolutionary pressures that turned them into *the* (partly, at least) *ir*rational animal. The latter is then a conclusion that BR shares with H&B.

4. Making sense

A striking aspect of this dispute is that all of the above reviewed standpoints use what look like reasonable evolutionary considerations in order to justify what are in fact contradictory views on human rationality. How can we make sense of this?

My diagnosis is the following: the different parties in the debate are implicitly using the term "rationality" in their arguments in more than one sense; but they seem not to be aware of the fact that they are doing so. Thus, the discussion gets muddled. In order to improve things, I will introduce an important conceptual clarification proposed by Kacelnik (2006).

4.1. *Rationalities*

Kacelnik's clarification proposal is the following: if we are going to discuss meaningfully whether or not non-human animals are rational, we need first to introduce a sharp distinction between three *different* concepts of rationality. These are discriminated below.

First, *P-rationality*, or psychological rationality; this is a concept of rationality associated with the presence of an articulated belief-desire network. Depending on what theory one accepts of what the nature of this network is, P-rationality may or may not be associated with procedural rationality. Theoreticians working within the H&B or ER views are clearly making this association.

Second, *E-rationality,* or economic rationality; this is a concept of rationality associated with behavioral consistency, revealed preferences and maximization of individual expected utility.

Third, *B-rationality*, or biological rationality; this is a concept of rationality associated with fitness maximization.

4.2. *Re-describing the debate*

Now, as mentioned above, I will use Kacelnik's three definitions above in order to reinterpret the ongoing debate in what I take to be a much more illuminating way. By this, I mean not only that this reinterpretation will make each of the positions and arguments come out clearer, but also that it will allow me to detect and isolate an important flaw afflicting all but one of the above reviewed approaches. Thus, it will make it possible for me to suggest a way of adjudicating between them. Let me try to show you this.

First, I will re-describe the main critical argument H&B or ER supporters put forth against the UR or OuC view. This re-description has the following aspect:

Premise 1: UR and OuC theoreticians contend that humans are (approximately) E-rational.

Premise 2: In order to be able to be, even approximately, E-rational, humans would need to be strongly P-rational.

Premise 3: It is both psychologically false and computationally and, therefore, biologically, impossible for them to be strongly P-rational.

Conclusion (by *Modus Tollens*): They cannot be E-rational either, not even approximately.

Second, I will re-describe the counter-argument UR or OuC supporters tend to advance against the criticism above according to the following *Modus Ponens* structure:

Premise 1: Natural creatures are B-rational.

Premise 2: In cognitively sophisticated natural creatures, B-rationality implies E-rationality.

Premise 3: Humans are cognitively sophisticated natural creatures.

Conclusion: Humans are E-rational.

Third, I will render Stanovich's argument in favor of BR as follows:

Premise 1: In Nature B-rationality is pervasive.

Premise 2: For non-human animals, being E-rational is easy.

Premise 3: For humans, being E-rational is very hard.

Premise 4: The worlds non-human animals live in and their cognitive systems are simple.

Premise 5: Human worlds and human cognition are highly complex.

Conclusion: Human cognitive P-complexity is the differentiating causal element that severs (in humans) the otherwise straightforward connection holding in Nature between B-rationality and E-rationality.

The clarity introduced by this re-description of the arguments pays off. For instance, it is now easy to see that, if Stanovich is right, then the two following conclusions will hold. First, strong P-rationality *cannot* be a necessary condition for E-rationality (*pace* H&B and ER theorists); under this assumption, the cognitively inspired objections these theorists raise to Friedman's or Anderson's account of human rationality lose their point. Second, human P-complexity is a *hindrance* to human E-rationality though (*pace* UR and OuC theorists); therefore, critics of the evolutionary justification for the use of classical rational choice theory as a theory of human behavior will nevertheless be right, albeit for the wrong reasons.

4.3. *Is the "Third Way" a better way?*

The question, of course, is: Is Stanovich right?

His argument is an inductive causal argument of the Millian sort (method of difference). It identifies human P-complexity as the differentiating causal element that accounts for the distinction allegedly observed between human and animal behavior regarding E-rationality. It thus assumes that, in the absence of this human peculiarity, E-rationality would follow, without further ado, from B-rationality.

As we have seen above, that E-rationality follows straightforwardly from B-rationality (in cognitively sophisticated creatures such as humans) is also a premise of the counter-argument by *Modus Ponens* UR or OuC supporters tend to put forth against Objection (B) (see section 2.2.).

But is it really the case that the connection holding between B-rationality and E-rationality is a straightforward one, either in humans or in non-human animals? This is the question that needs to be addressed in order to clarify whether or not the evolutionary justifications advanced by Stanovich, UR or OuC views are indeed adequate.

5. B-Rationality

In order to try to find out an answer to this question, let me take a closer look at the concept Kacelnik's term 'B-rationality' is meant to capture (Kacelnik 2006, pp. 93-96).

5.1. *Fitness Maximization*

As mentioned above, the concept of B-rationality is to be associated with the idea of fitness maximization. On the other hand, the idea of the fitness of a biological agent is associated with its degree of success relative to that of other biological agents in the same population. Fitness is thus a relative rather than an absolute notion.

But what does "success" mean for a biological agent? And, whatever it means, how is it measured?

The idea of "biological success" is clearly associated with survival and reproduction. However, complex organisms neither produce copies from themselves nor survive to themselves. As a matter of fact, what increases or decreases within a given population across long stretches of time is something else, namely, the number of copies of a gene relative to the total *loci* available for that gene in a given population. Thus, it is not organisms but genes that enjoy of more or less biological success. Assuming, *pace* Godfrey-Smith (Godfrey-Smith 2009), that we are allowed to import the agency idiom into evolutionary biology, genes *are* the biological agents (cp. Dawkins, R. 1976, 1982, Dennett, D. 1995).

5.2. *Inclusive Fitness*

Now, like everything else biological, the behavior of an individual organism is also, in some sense of this term, "guided" by the genes it carries. Thus, it should be possible, at least in principle, to analyze it as a derivative form of implementing fitness maximization. Kacelnik (2006) claims that this is precisely what theoretical population geneticists do.

According to him, the idea there is to work out a function describing the behavior of particular biological organisms as the behavior of maximizing agents; again, in order for this to be possible, organisms need to be seen as extensions of their genes. What is then being maximized is what theoretical population geneticists call *inclusive fitness*. And this is still to be understood at the genetic level (i.e., relative success of particular genes within the gene pool) rather than at the individual level.

However, putting into practice the guiding principle that B-rational individuals are supposed to exhibit a behavior maximizing their inclusive fitness is no easy task. Kacelnik lists a number of problems afflicting the use of this principle in empirical research. Some of them are the following.

First, no organism can be expected to maximize inclusive fitness under all possible circumstances; thus, the concept is useful only when relativized to a limited number of them.

Second, the detection that a strategy deployed by a particular organism is a form of maximizing its inclusive fitness must occur together with the identification of a typical situation that is or has been evolutionarily relevant for that organism; but this is easier said than done.

Third, when evolutionary biologists engage in optimality research, what they are able to test in practice is which of different strategies a particular organism

consistently prefers. But, depending on the assumptions they make, any of the strategies being tested may be taken to be a proxy for a form of maximization of inclusive fitness. Thus, the inference that the one that seems to be chosen by the organism is the one that indeed maximizes its inclusive fitness under the circumstances being tested seems to be obviously circular.

5.3. *If B-Rational, then E-Rational?*

Regardless of how difficult it is to test B-rationality effectively, might nevertheless E-rational behavior be derivable, at least in principle, from the constraints of B-rationality?

There is a sense in which it is possible to answer this question trivially in the affirmative. As Kacelnik put it (Kacelnik 2006, p. 99):

> if a subject is a consistent maximizer of inclusive fitness, then it is a consistent maximizer of something, and a consistent maximizer of anything is, by definition, E-rational.

But, in spite of the trivial proposition above, does it really follow that the assumption that an organism maximizes inclusive fitness readily translates into the existence of some particular form of E-rational consistency among its revealed preferences?

As a matter of fact, the answer seems to be no. In spite of Jensen, Call and Tomasello study and others like it, a look at the literature on animal rationality reveals that highly plausible suggestions have there been made on how to interpret in a B-rational consistent way E-irrational (i.e., inconsistent) preference rankings which have also been conclusively observed (see, e.g., Schuck-Paim, C., Pompilio, L. and Kacelnik, A. 2004 or Houston, McNamara and Steer 2007a, 2007b).

Truth be told, some of the very same authors mentioned above have also shown that with the introduction of suitable changes to the definition of either the relevant problem space or the utility function being used the exact same behavior could be reinterpreted in a way that would make preferences come out as E-consistent after all. The problem with implementing this sort of tinkering is, of course, that, once this is systematically done, E-rational interpretations of animal behavior begin to look more and more *ad hoc* and end up running the risk Kacelnik identified of becoming trivial.

The upshot of this work seems then to be the following: the difficulties brought about by an E-rational interpretation of animal behavior are *not essentially dissimilar* to those brought about by more familiar attempts to produce E-rational interpretations of *prima facie* E-inconsistent human behavior. I.e., in this respect human and non-human animals are not that different after all!

6. There is no human exceptionality

Introducing into this discussion the distinction between P-rationality, E-rationality and B-rationality along the lines suggested by Kacelnik has therefore proven to be valuable. It led us to the obtaining of the following important conclusion: it is not possible to straightforwardly derive a useful behavioral concept of E-rationality, both

for humans as well as for non-human animals, from the assumption that we are all subject to the constraints of B-rationality.

However, that such a derivation should be straightforward and unproblematic for humans is precisely what is taken for granted by the second premise of the evolutionary counterargument put forth by the UR or OuC approaches.

On the other hand, that such a straightforward and unproblematic derivation should be the default evolutionary assumption is something Stanovich's Millian argument takes for granted. It is the realization that such a default seems not to apply to human behavior that motivates his search for the odd causal factor responsible for this human exceptionality.

But, in reality, there simply is no human exceptionality in this regard. Birds, rats or chimpanzees don't do better than humans in terms of rational agency!

As a matter of fact, only under a limited and highly constrained set of circumstances does it seem to make sense to expect that B-rational and E-rational interpretations of human or non-human animal behavior will come out uncontroversially aligned. These are the circumstances under which maximization of individual expected utility may be taken to be an accurate proxy for inclusive fitness.

7. Where do we stand?

The evolutionary *rationale* claimed by the holders of the UR and OuC views as well as by Stanovich's BR approach relies on a false premise. It is, therefore, not acceptable.

In contrast, the H&B and the ER views seem to have clearly understood that assuming living creatures to be B-rational in no way legitimizes an analysis of their individual behaviors as being E-rational in any straightforward reading of this term.

However, H&B theoreticians accept that classical rational choice approaches to E-rationality *define* behavioral and cognitive rationality. Thus, they conclude that evolution made humans non-rational. This strikes me as wrong.

As a matter of fact, this use of "rational" and "rationality" by H&B theoreticians seems to me to still fall prey to a top down view of Man. That is, a view according to which Man would ultimately be a sort of creature left somewhere halfway between "divine" rationality and "animal" nature – a being endowed with "indirect" rationality only, if I'm allowed to use here the expression once coined by J. Elster (1984). But what I think we need in this respect is to replace, once and for all, such a view of Man with a genuinely naturalistic, bottom up, view.

This is exactly what, if I understand them correctly, ER theoreticians try to do. As already mentioned in 1.4. these researchers contend that the fast and frugal heuristics they claim to have identified in the human cognitive apparatus, rather than being simply dirty cognitive shortcuts, are frequently more robust ways of dealing with an uncertain and complex world than traditional optimization tools. To my view, if true, this contention changes substantially the nature of the philosophical debate concerning the understanding of human action.

Thus, the ER approach seems to me to be the most promising. However, in most cases, we simply do not know yet how to connect animal or human behavior and cognition with what Dennett calls 'strategic design' (another word for B-rationality really) or with P-complexity (assuming that it is correct to associate P-rationality with procedural rationality). In fact, an approach such as this is more like a research program than like a full-blown view of the issue.

To conclude, it seems to me that this is an area of research in which we need to be humble – what the putative future *empirical* successes of ER or other similar research programs might entail for our understanding of the specificities of human or animal decision-making behavior is something we simply cannot anticipate by means of purely analytic reasoning.

University of Lisbon
Departamento de Filosofia/
/Centro de Filosofia da Ciência

antoniozilhao@fl.ul.pt

Portugal

REFERENCES

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école americaine. *Econometrica* **21**, 503-546.

Anderson, J. R. (1990). *The Adaptive Character of Thought.* Hillsdale (NJ): Lawrence Erlbaum.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences* **14**, 471-517.

Arrow, K. (1951). *Social choice and individual values.* New Haven (CT) & New York (NY): J. Wiley, Chapman & Hall.

Chase, V.M., Hertwig, R. and Gigerenzer, G. (1998). Visions of Rationality. *Trends in Cognitive Science,* **2,** 206-14.

Davidson, D. (1980). Hempel on Explaining Action. In: *Essays on Actions and Events.* Oxford: Oxford University Press*,* 261-76.

Davidson, D. (1995). Could There Be a Science of Rationality? *International Journal of Philosophical Studies* **3**, 1-16.

Dawkins, R. (1976). *The Selfish Gene.* Oxford: Oxford University Press.

Dawkins, R. (1982). *The Extended Phenotype.* Oxford: Oxford University Press.

Dennett, D. (1995). *Darwin's Dangerous Idea – Evolution and the Meanings of Life*. London: The Penguin Books.

Elster, J. (1984). *Ulysses and the Sirens – Studies in Rationality and Irrationality.* Cambridge: Cambridge University Press.

Evans, J. St. B. T. (1991). Adaptive cognition: The question is how. In *Open Peer Commentary to J. R. Anderson's Is human cognition adaptive?* Behavioral and Brain Sciences **14**, 493-4.

Friedman, M. (1953). The Methodology of Positive Economics. In *Essays in Positive Economics.* Chicago (IL): The University of Chicago Press.

Gigerenzer, G. (2000). *Adaptive Thinking – Rationality in the Real World.* Oxford: Oxford University Press.

Gigerenzer, G., Todd, P. and The ABC Research Group (1999). *Simple Heuristics that Make us Smart.* Oxford: Oxford University Press.

Godfrey-Smith, P. (2009). *Darwinian Populations and Natural Selection.* Oxford: Oxford University Press.

Houston, A. I., McNamara, J. and Steer, M. (2007a). Do we expect natural selection to produce rational behaviour? *Philosophical Transactions of the Royal Society* **362**, 1541-1543.

Houston, A. I., McNamara, J. and Steer, M. (2007b). Violations of transitivity under fitness maximization. *Biology Letters* **3**, 365-367.

Jeffrey, R. (1983). *The Logic of Decision* (2nd rev. ed.). Chicago (IL): Chicago University Press.

Jensen, K., Call, J. and Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science* **318**, 107-109.

Kacelnik, A. (2006). Meanings of Rationality. In: Hurley, S. and Nudds, E. eds. *Rational Animals?* Oxford: Oxford University Press, 87-106.

Kagel, C. J. (1987). Economics according to the rats (and pigeons too): What have we learned and what we hope to learn. In Roth, A., ed. *Laboratory experimentation in economics: Six points of view.* Cambridge: Cambridge University Press, 587-703.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P. and Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Lichtenstein, S. and Slovic, P. eds. (2006). *The construction of preference.* Cambridge: Cambridge University Press.

Piatelli-Palmarini, M. (1994). *Inevitable Illusions: How Mistakes of Reason Rule Our Minds.* New York: Wiley & Sons.

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

Schuck-Paim, C. and Kacelnik, A. (2002). Rationality in risk-sensitive foraging choices by starlings. *Animal Behaviour,* **64**, 869-879.

Schuck-Paim, C., Pompilio, L. and Kacelnik, A. (2004). State-dependent decisions cause apparent violations of rationality in animal choice. *PLOS Biology,* **2**, 2305-2315.

Searle, J. (2001). *The Rationality of Action.* Cambridge (MA): The MIT Press.

Simon, H. (1957). *Models of Man - Social and Rational.* New York: Wiley & Sons.

Simon, H. (1969). *Sciences of the Artificial.* Cambridge (MA): The MIT Press.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin,* **119**, 3-22.

Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning* **19**, (1), 1-26.

Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy*, **69**, 213-225.

Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review,* **76**, 31-48.

Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis* **9**, 163-17.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* **185**, 1124-1131.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice,. *Science* **211**, 453-458.

von Neumann, J. and Morgenstern, O. (1944). *The Theory of Games and Economic Behavior.* Princeton (NJ): Princeton University Press.