Khalil, Elias L.

Working Paper
Charles Darwin meets Amoeba economicus: why natural selection cannot explain rationality

Papers on economics and evolution, No. 0622

Provided in Cooperation with:
Max Planck Institute of Economics

Suggested Citation: Khalil, Elias L. (2006) : Charles Darwin meets Amoeba economicus: why natural selection cannot explain rationality, Papers on economics and evolution, No. 0622, Max-Planck-Inst. für Ökonon, Jena

This Version is available at:
http://hdl.handle.net/10419/31838

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Charles Darwin meets Amoeba economicus: Why Natural Selection Cannot Explain Rationality

by

Elias L. Khalil
Charles Darwin meets *Amoeba economicus*: Why Natural Selection Cannot Explain Rationality

Elias L. Khalil

**ABSTRACT**

Advocates of natural selection usually regard rationality as redundant, i.e., as a mere linguistic device to describe natural selection. But this “Redundancy Thesis” faces the anomaly that rationality differs from natural selection. One solution is to conceive rationality as a trait selected by the neo-Darwinian mechanism of natural selection as . But this “Rationality-qua-Trait Thesis” faces a problem as well: Following neo-Darwinism, one cannot classify one allele of, e.g., eyesight as better than another without reference to constraints—while one can classify rationality as better than irrationality irrespective of constraints. Therefore, natural selection cannot be a trait. This leads us to the only solution: Rationality is actually a *method* that cannot be reduced to a trait. This “Rationality-qua-Method Thesis” lays the ground for alternative, developmental views of evolution.

**Key words:** Redundancy Thesis, rationality anomaly, Rationality-qua-Trait Thesis, incoherence problem, Rationality-qua-Method

**JEL Classification:** D0

---

1 Email: elias.khalil@buseco.monash.edu.au. Department of Economics, Monash University, Clayton, VIC 3800, Australia. The title of this paper is inspired by the title of Brian Skyrms’s [1994]. This paper was supported by the Konrad Lorenz Institute for Evolution and Cognition Research (Altenberg, Austria), the Max Planck Institute of Economics (Jena, Germany), and Monash University’s Faculty Research Grant Scheme, 2006. This paper, at different stages of its development, received comments from Richard Posner, Ulrich Witt, Gerhard Müller, Werner Callebaut, Steven Orzack, Steve Abedon, Jack Vromen, Brian Charlesworth, Gordon Tullock, Timothy Crippen, Michael Ghiselin, Howard Margolis, Robert Axelrod, Richard Levins, Richard Nelson, Joseph Lopreato, R. Preston McAfee, J.S. Metcalfe, Peter Taylor, Elliott Sober, Stanley Salthe, Casey Mulligan, Franz Weissing, participants of a seminar at the Konrad Lorenz Institute, and most recently Brian Skyrms. Also this paper benefited greatly from the assistance of Michael Dunstan. The usual caveat applies.
What does an amoeba do on an average day? It simply cannot afford to sit around and hope that some food will fall from the sky. It certainly cannot afford, because of the high cost of motion, to roam the neighborhood randomly in the hope of encountering some yeast to eat. The amoeba must take smart actions. When it moves in search of nutrients by using its “false feet,” it moves carefully towards higher food concentration gradients via chemotaxis. Also, it moves directionally away from toxic, unsuitable environments. However, when starved, these unicellular, solitary organisms have a strong incentive to undertake collective action. They form something equivalent to what economists call a “club” [Eichinger et al., 2005]. The solitary organisms congregate into a multicellular unit that appears as a slug. The homogeneous cell population becomes differentiated into different types of cells. The slug acts as a momentary “bus” that provides a more effective means of transportation to a new environment that will hopefully have a better concentration of food.

All organisms have to work, and they better work carefully as they search for nutrients, sexual partners, suitable habitats, and safety. All organisms must deliberate and choose among alternatives. The environment they face is neither uniform nor static. All organisms must make decisions, and make the best decisions possible, given the environmental incentives. If the making of the best decisions possible allows us to call human agents “Homo economicus,” it should equally allow us to call other living agents “Organism economicus” or, in the above case, “Amoeba economicus.”

But how could brainless organisms, from Amoeba economicus to plants, ever be rational? Rodolfo Llinás [2001] argues that plants do not need brains because they do not move. Brains, as the argument goes, are needed for animals that move in order for the animals to make predictions as they move. But there are a host of organisms, such as our Amoeba economicus, which move without the need for a brain. So, the brain might not be essential for rationality. But, first, what is rationality? As defined here, it amounts to two characteristics: Decisions must be consistent (i.e., avoid intransitive preferences) and decisive (i.e., avoid incompleteness).

With this minimal definition, plants make rational decisions with respect to the economic use of water, nutrients, and exposure to the sun. They make choices that are neither inconsistent nor indecisive. And they do not need a brain to make such choices. So, brains might be needed in complex organisms that have to coordinate many functions. They may not be simply the
product of mobility as suggested by Llinás. In simple organisms, there is no single specialized organ to take care of coordination. Instead, in brainless organisms, coordination of functions must be undertaken by less specialized tissues. If this is the case, we should thus not conflate rationality with the brain, as much as we should not conflate mobility with limbs or digestion with stomachs. Rationality, mobility, and digestion can be the prerogative of other, general-purpose parts of the body.

Can Darwin’s theory of natural selection explain rationality, which seems to be the characteristic of organisms as primitive as *Amoeba economicus*? The central thesis of the paper is that natural selection cannot explain rationality. Natural selection, by definition, is limited to the explanation of how a superior trait/technology becomes optimally spread in population. Given that rationality is not at par with traits/technology, it cannot explain rationality. The paper shows that the treatment of rationality as an optimized trait leads to logical incoherence.

Section 1 clarifies the discussion of “optimization” applies whether understood as bounded or unbounded. Section 2 demonstrates the uses of the term in economics and evolutionary biology. The rest of the paper, as shown in Figure 1, presents the structure of the central thesis.

**Redundancy Thesis:** *rationality anomaly* →
**Rationality-qua-Trait Thesis:** *incoherence problem* →
**Rationality-qua-Method Thesis:** *uncaused cause?*

*Figure 1: The Structure of the Argument*

Section 3 identifies the “Redundancy Thesis,” i.e., the idea that we do not need to discuss rationality since natural selection can substitute for rationality. Section 4, though, shows that the Redundancy Thesis faces the “rationality anomaly”, i.e., the fact that rationality optimization differs from natural selection optimization. A solution to this problem, as discussed in Section 5, could be that one should instead recognize rationality as a trait. Section 6, the core of the paper,
shows that the “Rationality-qua-Trait Thesis” runs into the “incoherence problem.” One form of the incoherence problem is that, according to neo-Darwinism or the theory of natural selection, all alleles of a trait, such as eyesight or beak shape, are equivalent. One cannot assert a ranking of alleles without reference to particular environmental conditions. On the other hand, one can define rationality as better than irrationality without reference to environmental conditions. So, the treatment of rationality as a trait leads to incoherence.

To avoid incoherence, Section 7 suggests that rationality is not a trait but rather a method to organize the employment and coordination of different traits. As a method, rationality cannot be the subject of evolution. This raises the issue about the origin of rationality. If rationality is not caused by evolution, what is the cause of evolution? Although this is an important question, it is not the focus of this essay. In any case, our inability to explain rationality should not lead us to rush and treat rationality as a trait.

The paper concludes by showing the ramifications of the “Rationality-qua-Method Thesis.” If rationality cannot be caused by natural selection, the natural selection paradigm has been oversold. This should open niches for alternative views of evolution. This includes, especially, the “evo-devo” (evolutionary developmental biology) approach [e.g., Müller & Newman, 2003]. It also includes the learning/developmental approach in economics as promoted by the Austrian school [e.g., Witt, 2003] and the Schumpeterian tradition [e.g., Schumpeter, 1989; Nelson & Winter, 1982; Dosi et al., 1988; Nooteboom, 2000].

This paper ignores empirical evidence, which sometimes does not support the selection story. While empirical and experimental studies are, of course, highly relevant [Orzack & Sober, 1994], this paper is not about whether the data support selection theory, but rather is about the logical coherence of the natural selection explanation of rationality. Also, the paper ignores many issues in evolutionary theory and rationality theory. For instance, it ignores the debate around the unit of evolution as being orthogonal to the definition of optimization as adaptation. The unit of selection has bedeviled the philosophy of biology [Wilson & Sober, 1989; Brandon & Burian, 1984; Williams, 1966; Sober, 1984], but, whatever is finally designated as being the highest unit of selection, the question of optimization will still persist.
1. Optimization: Unbounded or Bounded?

For rationality theory, optimization is the maximization of the utility function (or wealth). For selection theory, it is the maximization of the fitness function. Prior to the exploration of the differences between the two theories, what does optimization mean? As shown in Figure 2, there are two different views of optimization: the bounded and the unbounded. Amazingly, the two views appear in both selection theory and rationality theory. For unbounded selection theory [e.g., Dawkins, 1976, 1982], a population usually reaches the highest level of fitness, meaning that it will contain the largest or proper proportion of individuals of the superior type. For unbounded rationality theory, agents make the best decisions given the information available and the cost of further search. Both positions are disputed by the bounded view. For bounded selection theory, soft Darwinians such as Stephen Jay Gould [1977] argue that the outcome of natural selection is often haphazard and engenders ill-fitted organisms. This is the case because of stickiness, of different kinds, in nature that obstructs the full operation of selection. For bounded rationality theory, behavioral economists seized Herbert Simon’s [1957, 1977] notion of satisficing (to Simons’s great displeasure) to argue the following: given that deliberation, like searching for information, takes time and resources, the truly rational agent may adopt heuristics.
that are better, on average, than acting with unbounded rationality [see Gigerenzer, 2005, 2006]. So, they argue, if we take a broader account of all relevant costs of achieving the best solution, sometimes it is good to settle for the second-best. For instance, if a person calculates each time whether he takes the stairs or the elevator to his office he would, on average, lose more resources than if he takes on the habit of one repertoire.\footnote{In economics, some authors have also pointed out situations of market stickiness that resemble the stickiness than hinders evolutionary selection optimization pointed out by Gould. As a result of nonergodic feedbacks concerning innovation, market niche and other variables, the potentially more productive firms do not even appear [Khalil, 2000].}

To emphasize, the debates between the bounded and unbounded views of rationality optimization or selection optimization do not actually undermine the notion of optimization. Even when advances the bounded view, one must assume the optimization approach.

To illustrate, the debate of whether evolution can generate false beliefs is not a debate about optimization \textit{per se}, but rather about the possibility of bounded optimization [e.g., Stephens, 2001]. In fact, when theorists want to show the limits of optimization, they have to use optimization. That is, the observed or theorized limits of optimization (boundedness) are themselves optimal. For instance, Stephen Stich [1985, 1990] argues that suboptimal beliefs can be adaptive: in evolution, what matters is not finding truth, but finding beliefs that afford greater chances of survival and, hence, leaving greater offspring on average than other beliefs. If agents act very cautiously and develop inaccurate beliefs that engender risk-aversion that maximizes chances of survival, they can be favored by natural selection over agents who attend to the facts and develop more accurate beliefs that put them at greater risk. But such inaccurate beliefs do not indicate sub-optimality. They rather indicate the bounded nature of optimality.

Put differently, as suggested by bounded rationality, agents develop heuristics and shortcuts that economize on search. As a result, agents adopt inaccurate beliefs when more accurate ones cost more than expected payoff. Actually, as Werner Callebaut [1998, pp. 86-90] shows, the concern with fitness is one of the motivations behind Herbert Simon’s [1977] notion of satisficing: satisficing maintains that agents are interested in “sub-optimal”—in the sense of
inaccurate belief—insofar as they assure greater probability of survival/fitness than “optimal”—in the sense of accurate beliefs. Stephens [2001] calls such inaccurate beliefs “better-safe-than-sorry argument.” Such beliefs prompt agents to adopt risk-averse behavior when it appears that accurate beliefs would lead agents, at least in some occasions, to amass greater wealth or progeny. However, accurate beliefs, that prompt agents to be risk neutral, may actually, on average, decrease wealth or progeny if the greater accuracy entails costs unjustified by marginal benefits. So, risk aversion can be explained not as a taste, as economists usually do, but rather as a constraint to restrict recklessness [see Khalil, 2007b].

So, Stich [1985] misses the point: selection theory (and rationality theory) does not generate accurate beliefs—and selection (rationality) optimization understood as bounded should not generate accurate beliefs. Rather, they should generate inaccurate beliefs, such as risk-aversion, when the costs of extra accuracy are higher than expected marginal payoff. In this case, the inaccurate belief is optimal in the broader, bounded sense.

2. Optimization in Economics and Evolutionary Biology

I use the term optimization in the broad sense to: a) include either the bounded or the unbounded kinds of optimization; and b) denote maximization in selection theory (maximizing the frequency in population of the best trait/technology type) and in rationality theory (maximizing the objective of function of the agent).

As Figure 3 shows, selection and rationality theories are present in economics and evolutionary biology. It is the practice of economists to use rationality optimization, but it is less known that they also use selectionist optimization without often being fully aware of it. Likewise, it is the practice of evolutionary biologists to use selectionist optimization, but it is less known that they also use rationality optimization without full awareness as well.
Optimization criterion according to:

<table>
<thead>
<tr>
<th>Rationality Theories</th>
<th>Selection Theories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Economics</strong></td>
<td>Maximization of surplus/utility; theory of altruism</td>
</tr>
<tr>
<td>Minimization of effort per catch; inclusive fitness</td>
<td>Market selection of best technology and institutions</td>
</tr>
<tr>
<td><strong>Biology</strong></td>
<td>Natural selection of best genotype</td>
</tr>
</tbody>
</table>

*Figure 3: Optimization: Economics and Biology*

Whenever economists invoke forces of competition to discuss why certain technologies survive and become dominant over less adaptive technologies, they are actually using the selectionist argument. Even though there is no sexual or asexual replication in market

---

3 In fact, the invasion of natural selection arguments is not limited to economics. A growing number of social scientists are attempting to reformulate their respective disciplines after evolutionary biology. This is especially evident in psychology [Barkow et al., 1992], anthropology [Boyd & Richerson, 1985; Betzig, 1988], sociology [Machalek, 1992; Lopreato & Crippen, 1999], and political science [Masters & Gruter, 1992]. In economics (the focus in this paper), the invasion has taken diverse directions [Hirshleifer, 1982; Anderson et al., 1988; Witt, 1993; Hodgson, 1993, 2002; Nelson & Winter, 1982; Nelson, 1995; passim Koslowski, 1990]. A few thinkers even aspire to ground the first principles of biology on cost-benefit analysis [Ghiselin, 1992; Tullock, 1994]. And they may not be alone. The study of animal behavior,
competition, there is imitation. That is, imitation ensures replication. Imitation does no have to be intentional. The selectionist argument works even if we assume that imitation is random: The more successful type is bound to grow, via investment or consumption, and hence bound to become more conspicuous. It is more likely for all other agents to imitate the most noticeable trait. The imitation engenders the dominance of the most productive type. Consequently, the fittest or most successful type proliferates in the population, what is called here “evolutionary optimization” or what the literature calls “adaptation” or “fitness.”

On the other hand, whenever biologists invoke the notion of fitness in terms of maximization of surplus they are ultimately using the rationality argument. Ethologists and ecologists usually advance the thesis that organisms tend to act efficiently, i.e., maximize the output while assuming a given input or minimize the input while assuming a given output [e.g., such as the behavior of specific social insects, is based greatly on how agents exchange information and adjust behavior in light of cost and benefit [e.g., Detrain et al., 1999; Cassill, 2003]. It can even be concluded that the aspiration of a general theory of behavior is not unreasonable [see Knudsen, 2002].

4 The term “fitness” has many nuances—such as the actual versus the expected values in measures of fitness—which we may ignore. Such details are unrelated to our main argument [see Endler, 1986, pp. 27-51]. Furthermore, there is a slight ambiguity in the literature concerning the definition of fitness that arises in sexually reproducing organisms [Keller, 1987]. The measure of fitness in terms of the quantity of individuals born with the robust type differs from the number of individuals to which each fit agent gives birth for the simple reason that it takes two agents to replicate in sexual reproduction. A more important problem, which is overlooked here, is that natural selection in a sexually reproducing population may not necessarily engender fitness [Akin, 1979; Karlin & Lessard, 1986]. The selected differences at the phenotypic level may not be transmitted to the next generation because of the random reshuffling of genes which is responsible for the probabilistic character of Mendelian inheritance.
McFarland, 1977]. This argument is called here “rationality optimization” or what the literature calls “optimum foraging theory” [e.g., McFarland, 1977].

To classify optimum foraging theory as part of a generalized rationality optimization should not be surprising. This is because optimum foraging theorists borrowed their tools from rational decision making in economics [MacArthur & Pianka, 1966; Charnov, 1976]. Optimum foraging theory, to which some economists have contributed [e.g., Tullock, 1971], is about the organism's search/feeding behavior [see MacArthur & Pianka, 1966; Schoener, 1971; McFarland, 1977; Smith & Winterhalder, 1992; Winterhalder & Smith, 1981, 1992, 2000; Stephens & Krebs, 1986; Smith in Dupré, 1987]. The theory takes the organism's ability to extract information from the environment (such as day length, temperature, location of prey, and so on) as given by the trait/pattern type, and then analyzes the effective use of the respective type. According to the theory, animals, plants and other organisms choose the least costly method of producing or catching a given prey, nutrients, or sunlight, so that harvest per unit of time/effort is maximized. So, optimum foraging theory is identical to rationality optimization theory in economics: in both cases, the agent tries to optimize the objective function, given input endowment and capacity type.

Another line of research spearheaded by economists [e.g., Kagel et al., 1995] has shown how animals in laboratories, such as pigeons and rats, economize in their choices. This confirms the predictions of rationality optimization theory.

On the other hand, it may surprise a few to find that the inclusive fitness hypothesis is also classified as part of rationality optimization theory and, hence, is distinct from selection optimization theory. As elaborated elsewhere [Khalil, 2006], the hypothesis is identical to a major theory of altruism in economics. Both disciplines model altruism as the outcome of the efficient allocation (rationality optimization) of resources between the self and a significant other, in order to maximize a more inclusive function. In the case of economics, the function is the utility function. In the case of biology, the function is the fitness function.

3. The Redundancy Thesis

According to the Redundancy Thesis, whatever is produced by rationality, i.e., maximized
expected utility or maximized foraging output, is already accounted for by natural selection:

**The Redundancy Thesis:** We sometimes state that an agent makes a decision to maximize the objective function (fitness or utility) only as a shorthand to stating that the agent is programmed to make such a particular decision in reaction to particular circumstances—otherwise, the agent would decrease his or her optimal fitness and, on average, would be deleted from the population by the forces of natural selection.

There are many similarities between optimization in natural selection and optimization rationality. At least formally (mathematically), both involve the maximization of an objective function under constraints. Brian Skyrms draws on the formal similarity between the two kinds of optimization. He starts with the similarity, and only qualifies it under special circumstances:

The theory of rational deliberation and the theory of evolution both deal with processes which tend to move in the direction of a provisional optimum. In both areas, strategic interaction leads to complex game theoretic situations where the provisional optimum may be a moving target, and where equilibrium considerations must be introduced. In both disciplines, theories initially developed under simplifying independence assumptions need to be extended to deal with correlation [Skyrms, 1994, p. 503-504].

Skyrms shows that the two theories are extended differently in the case of correlation, i.e., when agents interact non-randomly with other agents in the population. So, as long as one assumes non-correlation, the two kinds of optimization are identical, at least formally.

The Redundancy Thesis is not merely about formal similarities. It amounts to material similarities as we will see in Elliott Sober’s [1998] defense below. It is also important to keep the Redundancy Thesis apart from the idea that rationality is a trait selected by natural selection. As shown below, the idea of rationality as being a trait is actually a response to the weakness of the Redundancy Thesis, which is that the Redundancy Thesis faces the rationality anomaly.

4. The Rationality Anomaly

The most important feature of neo-Darwinian theory [e.g., Mayr, 1976, 1982, 1988; Dawkins,
is that agents do not undergo change as a result of selectionist optimization. It is the population that undergoes improvement through "editing," i.e., the deletion of relatively less fit members. In contrast, the most important feature of rationality theory is that agents undergo change as a result of rationality optimization. What exactly is the difference in the conceptual apparatus of the two kinds of optimization? To answer the question, we need to distinguish two major criteria, the unit and the objective function as shown in Figure 4, with regard to each kind of optimization.\(^5\)

<table>
<thead>
<tr>
<th>Optimization according to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rationality Theory</td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Selection Theory</td>
</tr>
<tr>
<td>Utility or output</td>
</tr>
<tr>
<td>Frequency of trait/pattern</td>
</tr>
</tbody>
</table>

**Figure 4: Optimization: Rationality vs. Selection**

The unit criterion is the form that the objective function characterizes. For rationality optimization, the unit is the agent such as the cell, organism, colony of organisms, and human organization such as the firm. The agent makes decisions to satisfy its needs given its type and environmental constraints. For selection optimization, the unit is the population that is not organized as a colony or a group of cooperating agents. Thus, the population cannot be

\(^5\) Sober [1997] discusses an apparent third criterion which sets rationality optimization apart from selection optimization. Mindless organisms supposedly do not have subjective utility, but rather have the objective property of fitness. However, as shown earlier, utility optimization, which is used to characterize human decision making, parallels foraging optimization; and market selection parallels natural selection. Thus, although utility can be maximized by agents by anything with which they regard as conducive to their welfare, utility itself is still an objective property. Thus, Sober's third distinction is unwarranted.
considered a decision-making unit or, in Richard Dawkins's [1976] lexicon, a vehicle (even when the members of the population share the same taxonomic framework, or share, as what Dawkins calls, the same "replicator" [see Khalil, 1997]). A population in the neo-Darwinian sense is a unit consisting of members that interact for reproduction. An evolutionary change occurs when the members, for some exogenous shock, can be ranked according to differential levels of reproduction success—a ranking which the selecting mechanism (nature or market) can distinguish. A change that a population can experience is merely an unintentional product of the deletion of the less fit trait/pattern types as measured by the infrequency of their offspring. Such a change is not occasioned by an active decision as in the case with the change which one experiences when one acts rationally.

Concerning the objective function criterion, it is not identical for both kinds of optimization. For rationality optimization, the function is either the agent's utility or the agent's output or offspring. For selection optimization, the function is the frequency of the adapted trait/pattern type in a population. In selection, the selector (nature or the market) "makes" the population of organisms or firms dominated by the best trait/pattern type possible via-à-vis the environment.6

6 In the sense used here, selection should be distinguished from the account of the rise of conventions or standards which are usually welfare- or fitness-neutral vis-à-vis the environment. The stability of conventions—such as using the metric system or particular facial expressions to express disapproval—depends on what other members of the group are doing [Young, 1996]. In biology, the theory of "evolutionarily stable strategy" (ESS) and evolutionary game theories provide, inter alia, an account of conventions [Maynard Smith, 1978, 1982; Maynard Smith & Price, 1973; Vincent & Brown, 1988; Hammerstein & Selten, 1994]. In this role, an ESS is found to be stable if all members in the pertinent population adopt it, which makes the group immune from the invasion of other competing strategies. In contrast, the stability characterizing the fitness of a population to its environment is a substantive property, i.e., not conditioned on the unison of actions of members. This paper is concerned exclusively with substantive properties which are usually welfare- or fitness-sensitive vis-à-vis the environment.
Put succinctly, while the unit is a decision-making one in rationality optimization, it is not so in the case in selection optimization. Furthermore, while selection theory is concerned with explaining the frequency of types, this is not the concern of rationality optimization theory. Thus, the two kinds of optimization have different conceptual apparatuses. Biologists somewhat sense the difference between the two kinds of optimization when they label the function which the organism optimizes as the "fitness function," and the function which nature optimizes as the "adaptive or superior fitness function."

Elliot Sober [1996, 1998, in Orzack & Sober, 2001] discussed seriously the difference between rationality optimization and selectionist optimization. (As shown below, however, Sober backs down from the emphasis on the difference.) In selection optimization, there is no actor to start with:

Deliberation involves a change that occurs in an individual; evolution effects a change in the composition of a population, the individual members of which need never change their traits at all [Sober, 1998, p. 408].

So, given the fact that rationality involves an agent with a different kind of function when compared to selection, the Redundancy Thesis is fallacious.

5. The Rationality-qua-Trait Thesis

If one rejects the Redundancy Thesis, how is rationality related to natural selection? If the agent transforms itself and develops as it deliberates, how does such a development relate to selection optimization?

The usual answer to this issue is that rationality is a trait. The Rationality-qua-Trait Thesis seems straightforward. Even non-biologists, such as W. Quine [1969] and Jerry Fodor [1981, p. 121], point out that if humans are not rational, they would have long become extinct. And the same applies, apparently, to nonhuman animals. Some thinkers question whether all nonhuman animals are rational [Witt, 2003; see Khalil, 1996]. Nonetheless, whenever one observes rationality, one tends to postulate that it is the product of selection optimization.

There is a problem with the Rationality-qua-Trait Thesis, which is discussed below. In the meantime, the Thesis provides a room to show that rationality differs from selection
optimization. For one thing, the product of a mechanism cannot generate outcomes similar to the outcomes generated by the mechanism itself.

Sober [1998] disputes this observation. He argues that even if rationality-as-trait is the product of the mechanism of selection optimization, it generates outcomes that are identical to the outcomes generated by the selection optimization. If this is the case, the Redundancy Thesis can assert itself from the back window, and hence Sober’s argument needs a close scrutiny.

For Sober, the two kinds of optimization are, in the final analysis, identical because they generate the same outcome. This is what he calls the "heuristic of personification": if the organism is given the chance to choose a trait present in the population, it would choose the best trait, i.e., the most productive. For instance, if a zebra were given a choice, holding everything else constant, it would choose the trait of fast speed over slow speed given environmental conditions. For Sober, given that both kinds of optimization engender the same outcome, there is no substantial difference between the two mechanisms in the final analysis.

Sober's thought experiment about the rational choice of zebras correctly entails that the agent adopts and uses efficiently the best technology in light of the decision maker’s environmental constraints. Selection optimization, on the other hand, ensures the identification and the proliferation of the best trait in light of environmental constraints. Even if the environmental constraints facing both optimization mechanisms are identical, rationality optimization and selection optimization do not necessarily engender the same outcome, contrary to Sober’s argument. Agents could select diverse technologies in accordance with their diverse tastes. And all these different kinds of technologies would meet the condition of rationality optimization. In contrast, selection optimization would set the different agents against each other, and measure whose type would be better than the other, according to a common criterion, such as the number of offspring or the size of monetary profit. Such a common criterion ignores differences in individual tastes.

To put it differently, while rationality optimization might find diverse ray of optimum balances between monetary profit and, e.g., leisure, selection optimization ignore differences among agents and ensures an outcome according to a uniform yard stick. Of course, the yard stick can be as complicated as the environmental conditions. Nonetheless, the yard stick is
common, which is absent in rationality optimization.

Thus, rationality optimization differs from selection optimization. The redundancy Thesis cannot be allowed to enter from the back window. Nonetheless, the Rationality-qua-Trait Thesis faces another problem.

6. The Incoherence Problem

The Rationality-qua-Trait Thesis faces a problem of incoherence, which is multilayered. The three sub-sections below discuss three such layers. The first layer shows the incoherence problem while assuming that traits undergo progress, i.e., ranking of traits without reference to environmental conditions or incentives, what is called here “constraints.” The idea of progress is antithetical to either selection optimization or rationality optimization.

If one subscribes to selection optimization, one cannot identify progress or what is the best trait in the abstract, without reference to the constraints. Selection optimization simultaneously determines, in light of the constraints, what is the best type and, consequently, what must proliferate throughout the population if there were no exogenous obstacles.

Likewise, if one subscribes to rationality optimization, one cannot identify progress or what is the best technology in the abstract. That is, one cannot rank different production possibility frontiers, based on technological progress, without reference to the constraints. Rationality optimization simultaneously determines, in light of the constraints, what is the best type and, consequently, what is the efficient use of the resources if there were no imperfect information and other exogenous obstacles.

The second layer discusses the incoherence problem while assuming that traits can be ranked only in reference to environmental conditions—i.e., as conceived by selection optimization. The third layer discusses the incoherence problem while assuming that traits can be ranked only in reference to choices in light of incentives—i.e., as conceived by rationality optimization.

6.1 Assuming the Progress View

What does it mean to state that traits undergo evolution? Given that we are ignoring here
optimization, the term “evolution” here means progress. Progress entails that we can measure the improvement of the trait without regard to environmental conditions or incentives. Such conditions or incentives, called here constraints, are relevant only with regard to optimization—whether of the rationality or selection kind.

The criterion of progress may vary, depending on the trait under question. For instance, one can postulate that eyesight progresses according to the improvement of sharpness of image. For a vehicle, the criterion of progress is the speed of the vehicle. In either example, the trait becomes better as it approaches some ideal character defined in the abstract. The trait becomes better through mutation or invention. The invention/mutation allows the agent to increase its productivity in amassing greater surplus or offspring. The source of the invention/mutation can be random shocks or intention [Khalil, 2007a]. Whatever the source, it is irrelevant to the main issue. Namely, according to the progress view, we can rank the increase of productivity without regard to constraints.

From the progress view, the Rationality-qua-Trait Thesis entails that rationality, like all other traits, is subject to improvement in the abstract sense. This entails that there is no upper limit to the improvement of rationality—similar to the fact that there is no upper limit to the improvement of traits such as speed or sharpness of image. Even when particular tools, such as vehicle design or eye design, sets some limits, the limits can be surpassed by restructuring the technology or the trait itself. However, rationality, by definition, has an upper limit in the abstract sense. Even if we change the design of traits/technology, full rationality (100% rationality) does not change. So, there is an upper limit to the improvement of rationality, which undermines the idea that rationality is a trait. Even if evolution has not produced 100% rationality yet, there is no evolutionary change after reaching the telos. The same statement cannot apply to traits. Simply put, the progress framework cannot support the Rationality-qua-Trait Thesis.

6.2 Assuming the Selection Optimization View

Does the Rationality-qua-Trait Thesis fair better under the neo-Darwinian view than under the progress view? For neo-Darwinism, evolution is not about progress. It is rather about fitness or what is called here selection optimization. Fitness is the adaptation of a population to its
particular environment. When fitness increases, it means that the most appropriate trait, as measured in relation to environmental conditions, has proliferated throughout the population. The judgment of what is the appropriate trait and the judgment of what is the dominant trait in the population are the same judgment: both are determined by comparing the varied traits in the population according to their relative effectiveness in a particular set of constraints [Dawkins, 1976; Mayr, 1982]. For instance, if a long beak is the fittest given other available alleles and environmental conditions, a short beak can become the fittest under different conditions. It is impossible to state that the long beak, or the short beak, is the fittest in the abstract sense.

Likewise, within the temple of neo-Darwinism, we cannot state that 100% rationality is better than irrationality in the abstract sense. Both are simply different alleles of rationality and, hence, cannot be ranked without reference to particular constraints. So, it is possible that, under some conditions, for irrationality to be better than 100% rationality. We can, at best, call the different shapes of rationality by different names—similar to how we call the different shapes of beaks. All shapes of rationality are equivalent in the sense that they cannot be ranked innately, i.e., they can only be ranked with regard to particular constraints.

As discussed below, the issue of different shapes of rationality is unrelated to the issue of bounded and unbounded rationality. After all, bounded rationality is based on 100% rationality. To wit, the terms “100% rationality,” “irrationality,” and “bounded rationality” are expunged from the temple of neo-Darwinism because the terms imply some innate ranking.

At best, we can state that a particular shape of rationality, call it “Rj,” is the fittest shape under a particular set of constraints, call it Cj, while Rj is the fittest under Ci. We can suppose that Rj expresses the dominance of acts of addiction, such as gambling and alcoholism, while Ri denotes the dominance of courage ethics. If we encounter Rj under Cj, we can conclude that we have “bounded selection efficiency,” i.e., we have selection inefficiency in the sense that the selectors are imperfect. So, once selection inefficiency decreases, fitness would rise when, under Cj, Rj replaces Ri. That is, the population would attain greater fitness when it becomes dominated by alcoholism and other forms of addiction.

In this manner, addiction-qua-irrationality has no meaning in the house of neo-Darwinism. Addiction could be irrational only in the selection optimality sense, i.e., it would be irrational when the selectors are imperfect which results in suppressing the fitness of the
population from reaching its optimum. So, clinical doctors, true to neo-Darwinian stricture, should encourage their patients to undertake more irrational actions in some environmental conditions.

In this manner, and in light of selection optimization, the Rationality-qua-Trait Thesis cannot define rationality. This is paradoxical: The theory that aims to explain rationality cannot tell the difference between irrational acts such as addiction and rational acts. All are equivalent and can only be ranked in light of the constraints.

6.3 Assuming the Rationality Selection View

Let us assume that agents can select their own traits—as actually the case of humans acquiring skills to become carpenters or neurosurgeons. If rationality is a trait, such agents are also capable of choosing rationality.

To state that agents can choose a trait entails two assumptions. First, the chosen trait must be substitutable or fungible with other traits. If agents can only be carpenters, they have no choice. The fact that agents can make choices, the resources used to attain one product (such as tables) can be used to attain another product (such as logged timber). Second, the trait must be scarce. If it is free, such as air, there is no need to make a choice in the first place.

The two assumptions also underpin the Rationality-qua-Trait Thesis. So, rationality must be costly and fungible. Concerning fungibility, the agent decides between investment in resources to nurture rationality or investment in same resources to nurture, e.g., the skill of carpentry. Let us assume that the agent chooses greater skill in carpentry at the expense of rationality. Let us assume that after investing in carpentry, there was a totally unexpected shock. The expected return from rationality suddenly rises, while the expected return from carpentry remains the same. If the difference of returns is high enough, and the transaction cost of switching is low enough, the agent should forget the sunk cost of the earlier decision and start anew investment in more rationality at the expense of carpentry. But how could the agent make a rational decision of switching if he is already irrational?

More broadly, can an agent, starting with any degree of rationality (X), trade it with another trait (Y)? For such substitution to take place, he must use the very skill (X). This leads to self-contradiction. As an analogy, let us examine an agent who uses a tool such as a saw (X)
to make a product such as a table (Y). Can he give up X in order to gain more Y? If he gives up X, he would be forced to make less Y, *ceteris paribus*, for the very fact that he needs X to produce Y. That is, the making of tables (or making choices) requires a tool (rationality). And as such, one cannot substitute between the tool and the product. One can only substitute between one product (tables) and *another* product (logged timber) produced by the same tool.

Along the same reasoning, if rationality is a trait, one cannot use rationality-qua-tool to make decisions on whether to invest more on rationality-qua-trait at the expense of an alternative trait. The finding that the Rationality-qua-Trait Thesis is incoherent is about rationality *per se*, irrespective of whether one proposes that rationality is bounded or is unbounded. That is, the finding that choosing rationality is self-contradictory is orthogonal to the issue of adoption of short-cuts or heuristics that may lead, in particular instances, to sub-optimal ("irrationality") choices. As discussed earlier, agents adopt heuristics because of bounded rationality, i.e., to avoid expensive search or deliberation cost. The adoption of heuristics is not a decision about trading rationality *per se*, but rather is about whether to use calculation in each instance or to adopt a rule/habit given that calculation requires costly search and time-consuming calculation. Some thinkers [e.g., Stich, 1985; Stephens, 2001] confuse rationality *per se* issue and the issue of adoption of beliefs (heuristics). For agents to decide on the adoption of heuristics, they use rationality *per se*. And when they adopt general rules, which in some cases or in hindsight turn out fallacious, they are not giving up rationality *per se*. In fact, they retain rationality *per se* in order to adjust, give up, or adopt more heuristics.


The Rationality-qua-Trait Thesis leads to the incoherent world of self-reference of Escher. To avoid incoherence, we have to reject the Thesis. The alternative has been already suggested in the diagnosis of the incoherence problem. Namely, rationality can only be a tool or a method to make decisions because it cannot be an element (trait) about which decisions are made. In fact, rationality does not meet the first condition mentioned above concerning choice making. Namely, rationality is not fungible with traits. Also, rationality does not meet the second condition. Namely, rationality is not scarce.
While it has been shown that rationality is non-fungible with traits, we need to show why rationality is not scarce. To be scarce, rationality must be costly. The fact is that rationality is costless (which is, again, separate from the issue of heuristics in light of bounded rationality). The decision, e.g., to stop searching for information and to take action is costless. A prime example is the multitude of cases of moral hazard. When the government started to impose seat belt laws, for instance, economists predicted that agents would drive less cautiously, resulting in more accidents but less (thanks to the seat belt) fatalities per accident [Peltzman, 1975]. The tradeoff between safety and time gained by greater speed was never conscious. Also, agents who have insurance tend to adjust to the insurance, and adopt new, less-careful habits with regards to the locking of doors. Likewise, Brian Goff *et al.* [1997] show that baseball players in the American League adjusted their behavior in light of the 1973 change of the designated hitter rule.

To go back to Robinson Crusoe, he has to make a decision on whether to continue fishing or switch to hunting. So, he adopts a heuristic that states that, if he catches zero fish in the first two hours, then it is better to switch to hunting for the rest of the day. The establishment of this rule, like the search for resources, is costly. But the method used to determine the rule itself, i.e., rationality, is not costly. While the agent can economize on calculation, searching, and the use of other traits, he cannot economize on being economical. To make a choice implies that one is economical. Therefore, it is incoherent to assert that one can make a choice, or nature can make it on one’s behalf, on how much rationality to use.

As a method of making decisions, rationality is an endemic aspect of decisions. This “Rationality-qua-Method Thesis” entails that rationality is the way agents reconcile the objective function in light of the constraint function. And for the organism to do so, the method or the bridge between the two functions cannot be an element in either function. The method is rather about the use of non-method elements, such as eyesight, muscles, beak, and other tools needed for survival.

8. Conclusion

The argument of this paper developed in two stages. First, it refuted the Redundancy Thesis by
showing that rationality optimization differs from selection optimization. Irrespective of the origin of rationality, the objective function that rationality optimizes differs from the function that selection optimizes. Thus, rationality optimization and selection should not be assumed to be identical simply because they employ the optimization technique. Otherwise, one should also identify rationality with the physical principle of least action on the ground that both use the optimization technique.

Given that rationality and selection are different, what is the source of rationality? This question leads, as constructed here, to the second stage of the paper, viz., the Rationality-qua-Trait Thesis. It supposes that natural selection can explain rationality—namely, rationality is a trait like any other trait. The Thesis leads to incoherence. The paper illustrates the incoherence problems in three ways.

First, if rationality is a trait, it must be susceptible to improvements. Such improvements can approach greater perfection along an abstract scale of progress. But the ideal state need not be specified. This means that, similar to other traits, there is no upper limit to rationality. But rationality by definition cannot be improved beyond 100% rationality. This entails that rationality cannot be a trait. Second, if rationality is a trait, and complying with the neo-Darwinism stricture against an abstract scale of progress, one cannot rank different kinds of rationality. In fact, we cannot distinguish rationality from irrationality. All are different modules or flavors that can only be ranked in light of environmental conditions. But this contradicts the thesis that irrationality, such as addiction, is maladaptive in all circumstances. Third, if rationality is a trait, one can choose less of rationality to obtain another trait. If so, one must be using rationality as a tool to make such optimal allocation of resources. How could one, however, use rationality to make optimal decisions when it is already compromised and given up in favor of other traits?

If rationality is not the child of rational selection, this opens vistas to understand evolution beyond the paradigm of natural selection. This does not need to entail the invitation of medieval scholastics and the idea of “uncaused cause” as Darwinists [e.g., Dennett, 1995; Hodgson, 2002] have warned. It is true that Darwin undermined the metaphysics of divine intervention, but by postulating the importance of rationality, which is uncaused by natural selection, does not mean that it is beyond scientific study. It only means that such a study has
not begun, and in fact extinguished by the dominance of the natural selection paradigm.

On the other hand, the consequence of elevating the role of rationality to the prime stage has already begun. It is at the center of the evo-devo approach. While the term is new, early pioneers of evo-devo include Waddington [1953, 1957], Lørvup [1974], Gould [1977] and Stanley [1979; see Khalil, 1993]. More recent literature includes the work of Matsuda [1987], West-Eberhard [1989], Raff [1996; passim Müller & Newman, 2003] and Hall et al., [2004; see Callebaut & Rasskin-Gutman, 2005]. In economics, the learning/development approach can be traced back to Joseph Schumpeter [e.g., Nelson & Winter, 1982; Dosi et al., 1988; Nooteboom, 2000] and includes an Austrian strand [e.g., Witt, 2003].

The evo-devo literature amply shows the plasticity of the form of organism. The form is the result of ontogenesis (development) which is partially a function of the environment—a theme championed by Lamarckism. The Lamarckian theme should be music to ears of the advocates of rationality theory who also stress the relevance of the environment in influencing behavior, which economists call “incentives.” In fact, mainstream, neoclassical economics has been called the science of incentives in that it concerns how agents change their behavior in light of changes in the environment. Of course, Darwinian evolution also stresses the role of the environment, but only as a selection force, and not as an inducement for the agent to shape its own development. That is, for Darwinian evolution, the environment enters only ex post to validate or invalidate an invention. In contrast, for developmental evolution, the environment enters also ex ante to inform and influence the invention itself.
References


Skyrms, Brian. “Darwin Meets the Logic of Decision: Correlation in Evolutionary Game
Theory.” Philosophy of Science, 1994, 61, pp. 503-528.


Waddington, Conrad Hal. The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology,


