CrossMark

The behavioural ecology of irrational behaviours

Philippe Huneman[1] · Johannes Martens[2]

Abstract Natural selection is often envisaged as the ultimate cause of the apparent rationality exhibited by organisms in their specific habitat. Given the equivalence between selection and rationality as maximizing processes, one would indeed expect organisms to implement rational decision-makers. Yet, many violations of the clauses of rationality have been witnessed in various species such as starlings, hummingbirds, amoebas and honeybees. This paper attempts to interpret such discrepancies between economic rationality (defined by the main axioms of rational choice theory) and biological rationality (defined by natural selection). After having distinguished two kinds of rationality we introduce irrationality as a negation of economic rationality by biologically rational decision-makers. Focusing mainly on those instances of irrationalities that can be understood as exhibiting inconsistency in making choices, i.e. as non-conformity of a given behaviour to axioms such as transitivity or independence of irrelevant alternatives, we propose two possible families of Darwinian explanations that may account for these apparent irrationalities. First, we consider cases where natural selection may have been an indirect cause of irrationality. Second, we consider putative cases where violations of rationality axioms may have been directly favored by natural selection. Though the latter cases (prima facie) seem to clearly contradict our intuitive representation of natural selection as a process that maximizes fitness, we argue that they are actually unproblematic; for often, they can be redescribed as cases where no rationality axiom is violated, or as situations where no adaptive solution exists in the first place.

✉ Johannes Martens
martens.johannes@free.fr

Philippe Huneman
philippe.huneman@gmail.com

[1] CNRS/Paris I Sorbonne, Paris, France

[2] Université Catholique de Louvain (UCL), Brussels, Belgium

1 Introduction

In behavioural ecology, recent literature has increasingly been concerned with the rational or irrational behaviour of organisms (Houston et al. 2007, 2012). Thus, ecologists have been questioning whether hummingbirds make irrational food choices (Bateson et al. 2002), whether amoebas behave in a rational way (Latty and Beekman 2011) and whether honeybees rationally forage (Shafir 1994).

The wild rufous hummingbird *Selasphorus rufus*, for instance, eats from flowers that have different kinds of variance in nectar. Hurly and Oseen (1999) examined whether it chooses flowers with no variance in nectar, medium variance or high variance. When choosing between two flowers with different nectars, hummingbirds consistently prefer the flower with the lower level of variance. Yet, when choosing among three flowers with different nectars, hummingbirds prefer the flower with the intermediate level of variance. This violates what economists call "independence of irrelevant alternatives" (cf. Sect. 2), which is a mark of rationality.[1]

Strikingly, all those organisms biologists study do not have much cognitive capacities—at least not of the kind that are traditionally ascribed to rational agents in philosophy (Adams and Garrison 2013). So, from this latter perspective, the question arises as to how these claims about the rationality or irrationality of non-human individuals should be interpreted.

There are actually several ways of addressing this issue. First, one could (in a deflationary manner) analogize to a human agent in the same situation. Alternatively, one could argue that these claims are better understood according to what *the modeler* deems rational. Or, lastly, one could argue that these behaviours are indeed genuinely rational, because the concept of rationality is much larger than what we—as philosophers—usually think (i.e. by referring to beliefs, cognitive capacities and desires).

Whatever the chosen interpretation, the rationale behind these different studies seems to be a common assumption, which accepts natural selection as the main cause of the apparent rationality exhibited by the organisms in situ (foraging choices, mating strategies, predator avoidance, etc). This common assumption relies on the fact that both natural selection and rational choice behave as *maximizing processes*. Hence, just as natural selection favors traits or genotypes that maximize fitness, rational agents choose options that maximize their utility. Overall, this assumption seems quite intuitive. But how to understand it, and in which conditions it really holds, are still open questions. They are the ones that will be envisaged in the present paper.

The parallel between the economic behaviour of rational agents and the biological action of natural selection was formulated at the very beginning of

---

[1] The lack of cognitive abilities in many of the species studied does not prevent to talk about rational decisions, provided that rationality is defined in terms of behavioral outcomes rather than reasons for behaving.

behavioural ecology. Thus, in a seminal paper for foraging theory—a major field of behavioural ecology—MacArthur and Pianka noted:

> There is a close parallel between the development of theories in economics and population biology. In biology however the geometry of organisms and therefore environment plays a greater role. Different phenotypes have different abilities at harvesting resources, and the resources are distributed in a patchwork in three dimensions of the environment […] we undertake to determine in which patches a species would feed and which items would form its diets *if the species acted in the most economical fashion. Hopefully, natural selection will often have achieved such optimal allocation of time and energy expenditures,* but such "optimum theories" are hypotheses for testing rather than anything certain (MacArthur and Pianka 1966, our emphasis).

According to these authors, the choice of patches and diet made by an individual of a given species can therefore be expected to match what the members of this species *would* do "if they acted in the most economical fashion"—i.e. in a rational fashion.

Since this period, behavioural ecologists have made an extensive use of such optimality modeling, though mostly to explain traits whose reward in fitness do not depend upon what others are doing in the population. Evolutionary game theory was later designed to explain those cases where the behaviours of individuals cannot be understood independently of the behaviour of the other members in the population. In this latter case, the task consists mainly in determining the *evolutionary stable strategy* (ESS), namely the strategy such that, if every member of the population adopts it, no mutant strategy would be able to invade the population.

The difference between behavioural ecology and evolutionary game theory parallels the difference between the situations where individuals are playing alone against nature, and the situations where the behaviours of the individuals have a strategic component. In the first case, maximal expected utility predicts the choice, whereas in the second case what matters is the Nash equilibrium between choices of the players. Here, ESS is (roughly) the equivalent of Nash equilibria in the context of biological populations, except that, in the former case, the strategy is not "selected" by an individual, but maintained (at equilibrium) by the process of natural selection itself. That is why Maynard-Smith (1982) wrote in the introduction of his groundbreaking book that, in evolutionary game theory "the criterion of rationality is replaced by that of population dynamics and stability, and the criterion of self-interest by Darwinian fitness." (p. 2).

Given the relevance of these economical tools for behavioural ecology, it is certainly legitimate to ask whether organisms of a given species, when performing a certain variety of behavioural choices—selecting between plants to pollinate, habitats, possible mates, etc.—are *indeed* rational, in the sense that they would satisfy the formal criteria of rationality. According to McArthur and Pianka, as we have seen, they are expected to. But then, questions arise when they do not conform to those expectations: (i) how can we interpret the choice; and (ii) what does the choice tell us about the ascription of "rationality" or "irrationality" to non-human creatures, such as plants or animals.

It's not obvious whether all explanatory approaches used by behavioural ecologists and economists in this context are concerned with the same idea of rationality. In this paper therefore we will focus on whether the notion of "rationality" is equivocal or univocal. We will assess the antinomies between the rationality that natural selection allows us to expect and the irrationality in the economist's sense, by distinguishing two strategies likely to make sense of it, and characterizing them with respect to the role the concept of natural selection plays in each. In the second section, we will propose definitions of economic rationality and biological rationality, and emphasize some of their principled relations. More specifically, we will define "irrationality" as a discrepancy between rationality as it's expected by biologists according to the parallel between selection and rationality, and rationality in an economic sense. Then, in order to investigate the putative cases of natural selection for irrationality, we will distinguish between different explanatory strategies. The third section considers cases of irrationality where an apparent irrationality in behaviour could be understood differently if retraced to a distinct decision-rule (for instance by emphasizing the lack of information available in the context). The fourth section considers cases where natural selection seems to directly favor irrational behaviours. The last section shows that some of the alleged cases of direct selection in fact do not really involve irrationality, or a genuine contradiction between economic rationality and biological rationality.

2 Economic rationality and biological rationality

2.1 E-rationality

Rationality in economics is the maximization of a utility function, generally defined on the basis of the preferences of the agent (given her beliefs). Canonical formulations emphasize that a utility maximizer, once the utility function is defined, must respect the following clauses:

- Transitivity of preferences.
  If the agent prefers option A to option B, and option B to option C, then, facing A and C she will prefer A to C.
- Independence of irrelevant alternatives.
  If the agent prefers A to B, then she will not reverse her preferences when the options are A, B and C (she will chose either C over A and B, or A over B and C, but never B over A).

If one drops either (or both) of these premises, no maximization of expected utility is possible because either the maximizer of utility attested in pairwise contexts won't maximize utility in many-options contexts, or because she won't ensure maximization in enlarged choice contexts.

The status of these axioms, however, is often disputed. The most important question is whether all of the rationality axioms are a priori axioms. If so, they

cannot be inferred, which, in turn, questions whether they can correctly account for the observed behaviour of economic agents. Classical economists—facing the growing challenge raised by behavioural or psychological economics (e.g. Akerlof and Schiller 2009; Camerer 2003; Kahneman 2011) that real humans are not always behaving rationally in experiments—usually agree that rational axioms (such as transitivity) are, *as such*, a priori. Hence, whatever the rational agent is, the idea is that her choices should satisfy this condition—otherwise the very system of choices could not allow any proper modeling. In this respect, the a priori character of these axioms is precisely what confers to these axioms their normative character.

In a seminal attempt to synthesize the various threads of thoughts about rationality, Kacelnik (2006) distinguished economic rationality so defined (E-rationality) from both psychological-philosophical rationality (PP-rationality) and biological rationality (B-rationality). PP-rationality, on the one hand, is defined in terms of the consistency of the judgments and cognitive processes that lead to a conclusion and/or a decision—for instance, rational agents will use Bayesian updating procedures for their beliefs, be unlikely to fall prey to Dutch book arguments, etc. B-rationality, on the other hand, is defined by the fact that a behaviour maximizes the fitness of its bearer (or another biological maximand considered as a "proxy" of fitness). Kacelnik notes that PP-rationality cannot be ascribed to non-human agents, since it assumes strong cognitive capacities (and we do not know whether and how other organisms have cognitive capacities). E-Rationality, by contrast, is much more general, for its satisfaction depends on the actual behaviours of individuals and not on their justifications. In this respect, it is closer to the concept of B-Rationality, which is assessed by considering the actual fitness payoffs of the choices—notwithstanding the mechanism which produces them.

2.2 B-rationality

Utility and fitness—which correspond, respectively, to E-rationality and B-rationality—exhibit a number of important differences. First, utilities refer to subjective states of the agents, whereas fitness can be objectively measured (or at least, measured as a distribution of the offspring, such that its estimation does not raise more specific problems than determining objective probabilities based on frequencies). Second, the proper construction of a decision-making model in economics typically involves a formal representation of the beliefs of the agents, whereas in biology, this representation is only needed to the extent that organisms have evolved sensitivity to cues that co-vary with some relevant features of the world. And third, the utility function in economics is usually derived from the preferences of the agents, while in behavioural ecology, the preferences of the organisms are always derived from the fitness values of the phenotypic options—which, in turn, are determined by the very nature of the ecological pressures. Now, despite these differences, the parallel between biology and economics remains a close one; for it is easy to construe B-rationality as a particular instance of E-Rationality. This can be done, more precisely, by interpreting the utility function u(.) in terms of fitness

payoffs w, and the credence function in terms of objective probabilities (defined over those states that are ultimately relevant for the organism's fitness).

An important precision, however, needs to be made concerning the *level* at which the concept of rationality can be applied in biology. In effect, economics considers individual agents, whereas evolutionary biology focuses on populations (natural selection is a populational process). Yet, within populations, it is the individual organisms themselves that are making decisions; for it is the individual organisms that actually implement and manage the different resource allocations regarding the multiple aspects of their life-cycle (mating, habitat choice, life-history, and so on). Life history theory, for instance (Stearns 1992), is a whole research field concerned with the way organisms allocate their resources between reproduction at various moments of their life cycles. Thus, in this approach, organisms are B-rational in that they implement resource allocations in a distribution of investments across life cycles that maximize their fitness.

This duality of levels (populational vs. individual), specific to evolutionary biology, contrasts with the simplicity of economics, and suggests two interpretations of the equivalence between "rationality" and "selection" advanced by researchers in behavioural ecology.

First, this equivalence can focus on *natural selection* as a force driving population change towards adaptation—following the way population geneticists since Fisher and Wright thought of selection (Birch 2015). Here, we consider that the dominant trait in the population, when natural selection is at work, is the one that would be chosen when considering the population as an organism rationally selecting its phenotypes. Natural selection itself is thereby the rational decision-maker; or as Hammerstein (2002) notes, "the evolutionary process itself is conceived as the maximizing agent" (p. 72). When one attempts to explain the size of leaves, or the life expectancy or reproduction time proper to a species, for instance, this way of conceiving of natural selection makes perfect sense.

Second, the equivalence could hold between *organisms'* individual behaviour, and rational behaviour. This is the proper context in which the question of putative selection of irrational behaviours needs to be framed, and also where it is the most problematic. The rationale supporting this interpretation is that, when organisms have been partly shaped by natural selection, one can expect that their decision-making capacities—whatever the meaning of "decision", i.e. whether it is a real process or a specific projection of the modeler—are designed to be at least minimally reliable. Hence one can expect that their decisions will satisfy at least some criterion of rationality.[2] In this latter case, however, we need to assume phenotypic plasticity—understood as the ability to adjust one's behaviour to some relevant aspects of one's environment. For, surely, an organism deprived of any form of agency wouldn't be properly envisaged as "rational" (or "irrational") in the

---

[2] In this latter case, however, we need to assume phenotypic plasticity,—understood as the ability to adjust one's behaviour to some relevant aspects of one's environment. For, surely, an organism deprived of any form of agency wouldn't be properly envisaged as "rational" (or "irrational") in the first place. Phenotypic plasticity is actually pervasive among animals and plants (West-Eberhard 2005), so this assumption should not be (presumably) very onerous.

first place. Phenotypic plasticity is actually pervasive among animals and plants (West-Eberhard 2005), so this assumption should not be (presumably) very onerous.

It is under this second interpretation that (putative) irrational behaviours of some animals—namely failures of satisfying clauses such as the transitivity axiom or the independence of irrelevant alternatives axiom—appears problematic. For against our expectations, what they indicate is that we cannot always theoretically assume that organisms are designed to be E-rational (i.e. maximize fitness as utility) when their behaviour results from natural selection (B-optimality)—and precisely for their originating in natural selection.[3]

Before considering irrationality in behavioural ecology, one might wonder about the relation between those two (economic and biological) rationalities and the psychological rationality mentioned above, which is often more familiar to philosophers. As Kacelnik emphasizes, PP-rationality concerns the process of choice and not its outcome, and it makes perfect sense that in many cases natural selection relied on non-rational processes such as emotions, mimesis or faith, rather than rational deliberation, in order to reach some fitness-maximizing outcome (the field of mate search in primates is an obvious example of this fact). But PP-rationality becomes relevant from either an economic or a biological perspective if one is interested in explaining the behaviours of individuals *in the face of new information*, for in this context different rules of updating (Bayes rule, etc.) can have different epistemic values according to either the criterion of utility or fitness maximization. In this case, however, the emphasis is merely put on the consequences associated with the use of such or such inference rule, rather than on the process (psychological or not) that implements the rule. Hence, to the extent that we want to understand the link between E-rationality and B-rationality, focusing on PP-rationality per se is not directly relevant, for it is always possible to evaluate the value of a given rule using either utility or fitness as a criterion of choice.

2.3 The problem of the gap between E-rationality and B-rationality

In the remainder of this paper, we will review some of the main hypotheses that have been proposed to account for the existence of apparent violations of E-Rationality in the living world, as exemplified by the studies cited in the introduction.

Suppose that, given a set of choice situations, organisms in a particular species are such that their behaviour appears as E-irrational. How, then should we interpret their breaking the clauses of rationality?

First, it might be that the whole decision-making device is in fact not a product of natural selection. After all, it is well-known from the different branches of biology that natural selection can only act on *feasible* strategies; hence the set of strategies is constrained by what is developmentally possible: "the specification of the set of

---

[3] Notice that behavioural ecologists do not *directly* track the fitness function of organisms when they study a given behaviour. They select a variable that is arguably a good proxy for fitness: it can be net energy intake, gross energy intake, in some cases metabolism rates, or more general measures such as inclusive fitness (Davies et al. 2012).

possible phenotypes from among which the optimum is to be found (…) is identical to a description of developmental constraints". (Maynard-Smith 1982, p. 5). Accordingly, the set of possible strategies can be very far from optimizing strategies, and, in such cases, we can't expect decision-making mechanisms to be fitness-maximizing, precisely because they have not been designed to maximize fitness. Hence, if one were to model organisms' decisions as rational choices here, one would wrongly assume the maximizing action of natural selection—such a misconception would precisely fall under the critique famously made by Gould and Lewontin (1979) against the adaptationist method.

Second, the decision-making mechanisms could be the proper result of natural selection, but in a context where the genetic and/or ecological structure of the population prevents the realization of the optimal decision device. An obvious case of this divorce between selection and optimality is heterozygote superiority, since in this case, by definition, natural selection cannot drive the heterozygote phenotype to fixation, meaning that the population fitness realized through natural selection cannot be such that the population is composed of the highest-fitness individuals only. Moran (1964) first designed cases of genetic structure of populations in which, without constraints and drift, natural selection cannot achieve any fitness maxima. In this case, the organisms cannot be expected to make choices that maximize a utility function based on a proxy for fitness as their design is not a fitness maximizer. Thus, violations of rationality axioms should be expected.

In these two kinds of situations, however, E-irrationality is *not* a genuine problem because organisms are not B-rational in the first place—either because of the weakness of natural selection, constraints or drift, or because selection was not actually maximizing anything. In other words, the failure of E-rationality is not really problematic according to these scenarios because there is ultimately no reason to expect the organism's decision-making devices to be E-rational.

Now, we may also encounter cases where, while natural selection is arguably the cause of the behavioural mechanisms, the choices made by organisms in some range of environments are *not* by themselves satisfying all criteria for E-rationality. The example of the hummingbirds breaking the independence of added alternatives when they face a triple options set (Hurly and Oseen 1999) exemplifies this situation. In order to avoid the contradiction between natural selection as responsible for decision-making devices and such E-irrationality of the corresponding decisions (indeed, natural selection is supposed to yield B-rational devices that take E-rational decisions), two approaches appear possible. First, one can redescribe the putative irrationality so that the violation of rationality clauses disappears; or inversely, one can take it at face value. In this latter case, one can either view E-irrationality as an indirect consequence of B-rationality or demonstrate that selection directly yielded such a feature. The next Sect. 3 asks how selection can be indirectly responsible of irrationality, while the following Sects. 4 and 5 focuses on the way selection could directly lead to E-irrationality.

3 Cases where natural selection is an indirect cause of E-irrational behaviours

Witnessing irrational choices in behaviour is something behavioural economics is familiar with, starting with Kahneman and Tversky's experiments and their "heuristics and biases" program (Kahneman et al. 1982). Here, agents have been regularly shown to violate PP-rationality (like in the famous case of Linda-the-bank-employee (Tversky and Kahneman 1982) or E-rationality: for instance when buying a TV screen, the agents will prefer the $150 to the $200 TV screen, but, adding a $500 option, they will chose the $200 screen, which violates the axiom of independence towards additional options (see Tversky and Shafir 1992).

Some evolutionary psychologists tried to answer this challenge by explaining how those systematic irrationalities can be the result of evolution by natural selection. For example Cosmides and Tooby (1997) argued that ordinary failure at the Wason selection task (a task where the subjects are asked to select evidence for testing an abstract conditional rule) is attenuated by presenting the terms in the language of social contract, which suggests that people actually evolved to detect breaches in social contract *even though* those competences lead them to make wrong inferences in purely logical settings. Thinking in those terms, interestingly, exemplifies a typical explanation of irrationality which behavioural ecologists can use to view putative cases of irrationality in their domain as instantiating a difference in decision contexts between the biological explanation and the economical assessment of animals' decision-making. We detail it (3.1) and afterwards offer another strategy to interpret this difference, here labeled the "trade-off" strategy (3.2).

3.1 Mismatch hypotheses

This hypothesis questions the role of the environment in the evolution of certain specific competences (foraging or mating decisions, inferences skills, etc.). The current contexts in which they are applied (and fail) clearly differs from the context in which they were selected for: the Pleistocene. Evolutionary psychology is thus full of textbook examples, such as the fear of spiders, which may have been adaptive in the past but no longer fulfills this function.

In the cases of non-human animals displaying irrational features, an explanation could point out this kind of mismatch between a decision rule that was selected—hence biologically rational—and therefore economically rational in a past context (since it was maximizing a fitness-based utility function), but appears now to be maladaptive, fitness non-maximizing, and therefore E-irrational in present-day contexts. As Hammerstein writes,

> some of what is discussed in economics as bounded rationality [a rationality affected by biases] may be caused by the imperfection of components of the behaviour generating systems—components that work synergistically in the natural habitat. In an atypical habitat [such as modern-day environments, for

humans], the interplay of the same components may turn into a disaster. (Hammerstein 2002, p. 77)

Animals that do not respect the insensibility to added alternatives may be explained in this fashion: since they were not exposed, in their evolutionary history, to ternary choices but only to binary choices, they do not have decision-making devices allowing them to choose in the case of added alternatives. The example of hummingbirds cited above could be approached in such a way (Schuck-Paim 2003; Kacelnik 2006). Therefore their behaviour appears as irrational.

3.2 Trade-off hypotheses

Invoking a mismatch between the contexts of evolution by natural selection and the contexts where irrationality is typically expressed is not the only way to make sense of irrationality from a Darwinian perspective. The team of Gerd Gigerenzer has forcibly argued in the 2000s that the dimensions along which natural selection operates are not exactly the dimensions along which agent's choices are measured by economic criteria (Gigerenzer et al. 1999). Thus, while natural selection will favor behaviours that maximize fitness in the specific environments in which the organisms live, the environments that are *less likely* to be encountered by organisms are much less relevant for selection. From the viewpoint of selection, the decision-making devices are, so to speak, "allowed" to make *in these environment* choices that are not maximizing a fitness-based utility function. Thus, as long as the decisions made are efficient in terms of overall fitness gains, it does not matter if they are sometimes inaccurate.

To take a well-known case studied by Gigerenzer (1998), humans are not good at reasoning in terms of probabilities, but much better when problems are translated in terms of frequencies. Those researchers' explanation credits natural selection with being quite good at shaping minds to calculate frequencies, which is very useful given their environment. Of course, probabilities are more efficient when it comes to large numbers, but when the everyday environment does not often involve large numbers, frequencies offer greater practicality and less cognitive complexity, which, in turn, imposes less cognitive and developmental costs.

Natural selection, therefore, balances efficiency and accuracy: wholly accurate decision-making processes may be less efficient in the majority of selective environment considered, i.e. be more costly and not favored for selection. Typically, if a decision procedure implemented in an organism is such that most of the time it makes good choices (from a fitness viewpoint), and is less costly (e.g. in metabolism, in energy intake, etc.), or quicker than a procedure which never breaks the rationality clauses, then it will be selected, even if it may occasionally break those clauses and therefore does not realize a perfect fitness maximizer. Those outbursts of irrationality are a by-product of the decision-making protocol, which is indeed B-rational.

Gigerenzer calls those choice-protocols "fast and frugal heuristics", and the decision-maker guided by those heuristics "ecologically rational".[4] In this sense,

---

[4] Ecological rationality corresponds in biology to what behavioural economists call "bounded rationality" after Herbert Simon.

behavior can still be considered (ecologically) rational, dissolving the contradiction that threatened us.[5]

The recent "error management theory" developed by Martie Haselton builds on an analogous idea: it is a better solution for an organism (which is therefore favored by selection) to systematically make the errors that are less fitness-costly than to be always approximately accurate but risk making the most costly errors. The rationale behind the theory is that from the viewpoint of fitness and survival, the two errors (e.g. overestimating or underestimating a value) are not symmetrical, even if both are logically equivalent as errors. Nesse (2005) famously compared this to the braking distance of a car: overestimating and underestimating this distance are not symmetrical errors from the viewpoint of the driver's life. It is similarly better to overestimate a threat (react without an adequate reason) than underestimate it (avoid reaction, and then get wounded or killed). This can allow for PP-irrational beliefs in humans (for example, overestimating the seriousness of an epidemics, or overestimating—for a man—one's sexual attractiveness (Haselton and Buss 2000, 2009).[6] It can also allow for E-irrational decisions—notwithstanding the fact that the decision-making device is itself B-rational (i.e. understood as fitness maximizing).

The latter hypothesis, however, is vulnerable to the same criticism raised by Gould and Lewontin (1979) against adaptationism. Indeed, systematic rationalizations of irrational behaviours based on the assumption of B-optimality would probably be dismissed by these authors as mere "ad hoc explanation"—at least until quantitative models and empirical evidence can back up the hypothesis. Otherwise it is always possible to consider that a set of E-irrational decisions is the cost to be paid for gaining fitness maximization in the most likely contexts. Since counting the possible environmental contexts and weighting their probabilities is not straightforward, someone can easily invoke a specific metric on those possible contexts to justify her ad hoc hypothesis, which is exactly what the anti-adaptationist critique targets.

Two things should be noted here. First, the difference between those two explanatory strategies may not be easy to make in practice. In the "trade-off" approach, organisms have choice strategies that are realized in most of the environments they meet such that they ensure better fitness, but *at the cost* of errors, maladaptation and a decrease in fitness in other, *less likely*, possible environments. For instance those possible environments include environments where a modeler requires the subject to compute with probability values, or environments where there are no possibility of making the most fitness-costly kind of error, etc. In the "mismatch" approach, they have choice strategies that were fitness-maximizing in

---

[5] Far from contrasting with B-rationality, ecological rationality actually encompasses it: it is underdetermined given that it centers on a utility function defined by a task on which the modeler focuses; and when this task is fitness, it coincides with B-rationality.

[6] Error management theory explains why people avoid drinking in sterile glasses that are used to contain urine samples—as a clear case of overestimating a threat. A prediction would be that, when people prefer, let's say, green glasses over red glasses, when faced with ternary choice that includes urine sample glass, they would chose red or green glasses stochastically ("any glass is better than the urine glass" being the underlying rationale) and show preference reversal in the face of added independent alternatives.

past environments, but not in modern-day environments. Yet modern-day contexts could precisely be seen as contexts in which environments that were less likely are now more likely (e.g. computing with large numbers and the need for skills in probability calculus). So rather than seeing those approaches as two very different cases of discrepancy between B-rationality and E-rationality, one should really see them as two poles in a *continuum* of explanations.

4 Cases where B-rational seems to conflict with E-rationality

A common feature of the evolutionary approaches mentioned above (mismatch and trade off hypotheses) is that they represent natural selection as an *indirect* cause of irrational behaviour—that is, according to these hypotheses, being E-irrational is not *per se* adaptive. But at this point, one could also ask whether natural selection might not act as a *direct* cause of some E-irrational behaviours. In this section, we explore this (intriguing) possibility, by focusing on two representative examples. The first is a simple model designed by McNamara et al. (2014) showing that, when the current composition of the choice set affects the future (reproductive) expectations of the individuals, violations of E-rationality can occur. The second is a study by Okasha (2011) who shows that, in some risky environments, organisms can be selected to maximize some form of non-expected utility. A critical discussion of these two cases is provided in Sect. 5.

4.1 E-rationality may conflict with (B-rational) future reproductive expectations

In McNamara et al. (2014) model, a foraging individual faces a choice between two or three sources of food, labelled A, B and C. Each of these food items has an energy content e_X and a handling time h_X, which means that, if item X is chosen, the individual won't be able to choose another food item before h_X is elapsed. In this model, the absolute profitability of each item is measured by its ratio e_X/h_X, and serves as a proxy for its reproductive value w.

As usual in behavioural ecology, the individual is supposed to maximize its long-term rate of energy gain, so that, at any time, its choice is always B-rational. But according to McNamara et al. its choice won't be necessarily E-rational—not, at least, if the set of available options is allowed to *vary* over time.

To make their case, McNamara et al. assume that each item has a probability of disappearing and of (re)appearing at any time in the future. This probability is supposed to be independent of their consumption, so there is no "depletion effect" involved in the disappearance of an option. McNamara et al. also envisage different values for these probabilities, but, for the sake of expository convenience, we will assume that, after one unit of time, A and C always disappear from the choice set (whether or not they are chosen in the first place) while B remains available at any time in the future. Importantly, we will also assume that an option that is *not* initially available has a zero probability of appearing in the future.

Table 1 Absolute profitabilities of options A, B and C

Options	Energy content e	Handling time h	Profitability e/h
A	20	20	1
B	8	5	1.6
C	18	10	1.8

As can be seen from the values in Table 1, the absolute profitability of each food item is such that $e_A/h_A < e_B/h_B < e_C/h_C$. Yet, because there might be an opportunity cost associated with choosing the most profitable option—which is *not* represented by its absolute profitability—this particular ordering does not provide a correct basis for deriving the B-rational choice. Instead, the B-rational choice depends on future expectations, as can be seen from the following conditionals:

(i)   If A and B are both available, then B is the B-rational choice. Indeed, choosing A provides 20 units of energy for the next 20 units of time, whereas B, if chosen first, can be chosen three more times in the same interval to obtain a total of 32 units of energy. (Remember that B is still available once an item has been consumed).

(ii) If B and C are both available, then choosing C is the B-rational choice, for C yields greater energy unit per time than (twice) B, and B can still be chosen twice after the handling of C is over.

(iii) If C and A are both available, then A is the B-rational choice; for B is no longer available after C has been chosen for the first 10 units of time.

(iv) If A, B and C are simultaneously available, then C becomes the optimal choice, for B can be chosen twice *after* C has been chosen first, which gives a total of 34 energy units for the 20 time units interval.

From there, it is easy to see that B-rationality conflicts with E-rationality. Indeed, given (i), (ii) and (iii), and assuming B-rationality, it follows that A > C > B > A. So transitivity is violated. Furthermore, given (iii) and (iv), it follows that adding B to the choice set {A, C} "reverses" the individual's preferences between A and C—even though C is still preferred to B in the pairwise choice between C and B (without A). So the independence of irrelevant alternative axiom is violated too.

To explain these violations, McNamara et al. (2014) stress that the current availability of an item *impinges on* the future reproductive expectations of the individual, making its absolute profitability irrelevant as a criterion of B-rationality.[7] Thus, the initial composition of the choice set provides some crucial information about which options are likely to be available *in the future*, and these future options, in turn, influence the rate of energy gain of the individual—which accounts for the violations of both transitivity and the independence of irrelevant alternatives.

---

[7] For related studies, see Houston (1997).

In their paper, McNamara et al. write that "it can be adaptive for these principles to be broken" (p. 3). This suggests that the violation of the rationality clauses is not a by-product of some other (adaptive) behaviour, but is *itself* the selected behaviour. Yet, as we will see in Sect. 5, there is actually an alternative description of this case where the validity of these axioms is preserved.

4.2 The B-rationality of non-expected utility

A central principle of E-rationality in the face of risk is the principle of expected utility maximization (von Neumann and Morgenstern 1944). According to this principle, a rational agent facing a choice between uncertain outcomes (lotteries) should, first, determine the objective, marginal probabilities $p_1, p_2, \ldots p_n$ associated with each possible outcome $x_1, x_2, \ldots x_n$, and then choose the action that maximizes its expected utility across these different branches. Formally, the expected utility eu of an action corresponds to the *arithmetic* sum of the utility associated with each outcome, weighted by their marginal probabilities:

$$eu = \sum_i \mathrm{u}(x_i)p_i$$

In what follows, however, we present a biological example (Okasha 2011) where the maximization of long-term reproductive success (B-rationality) leads to an apparent violation of this principle.

Okasha imagines the following scenario. Suppose a foraging organism must choose between two options, labelled A and B. Choosing option A provides the individual with 5 energy units for sure, while choosing B provides the individual with either 9 energy units or 1 energy unit, with a probability of 0.5 each. Suppose then that, though both options A and B have the *same* expected energetic value (5 units), the organism ends up choosing the sure option A over the risky option B. How can we account for this choice, *assuming B-rationality*?
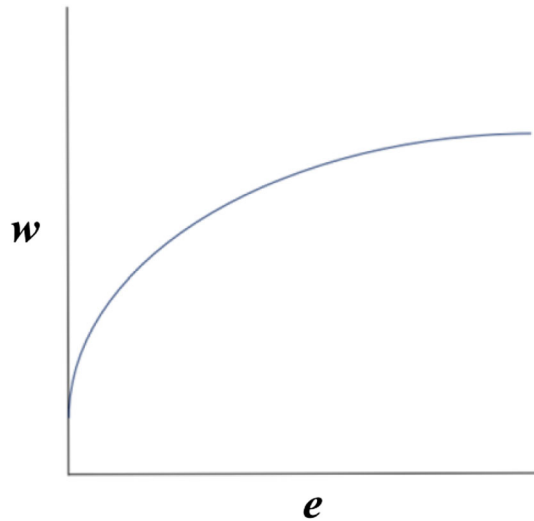
In accordance with expected utility theory, a possible explanation could lie in the *shape* of the function relating absolute fitness w to energy content e. Suppose, indeed, that there exists a concave relation between absolute fitness and energy content (Fig. 1).

Suppose also that biological utility u(.) is interpreted in terms of absolute fitness w. Accordingly, it would follow that $u(5) > p.u(1) + (1-p).u(9)$ for any probability p, so that the risk averse (B-rational) option would also be the one which maximizes expected utility.

Now, as pointed by Okasha, the concavity of the fitness function (diminishing marginal returns) is not the only possible explanation for the existence of B-rational risk averse behaviours; indeed, another possible explanation concerns the *nature of the risk* faced by the organisms (Robson 1996).

In the above example, two possible outcomes are associated with the risky option B, namely one in which the organism receives 9 energy units (the "good" outcome) and one in which it receives 1 energy unit (the "bad" outcome). Both, as we have seen, have the same probability 0.5 of occurring. Yet, each of these outcomes can be

realized in two different ways. On the one hand, their realization might be independent for each individual adopting the risky strategy (B). In this case, the risk is called *idiosyncratic*, for it is "as if" the realization of these outcomes was determined by a separate coin's flip for each individual (thus, if two individuals choose B and the first gets the good outcome, the second has still a probability 0.5 of getting the bad outcome). On the other hand, the realization of these outcomes might be correlated between the individuals adopting the risky strategy. In this case, the risk is called *aggregate*, for all of the individuals opting for B have a probability greater than 0.5 of ending in the same realized state—in the case of probability 1, it is "as if" the realization of the outcome were determined by a single coin flip for all the individuals (if two individuals choose B and the first gets, say, the good outcome, the second will get the good outcome with probability 1).[8]

As noted by Okasha, the nature of the risk (idiosyncratic or aggregate) does not affect the probability distribution faced individually by the organisms. Nonetheless, it has critical consequences for the link between B-rationality and E-rationality. Thus, if the risk is purely idiosyncratic, the principle of expected utility maximization (E-rationality) is automatically satisfied. In such a case, the long-term evolutionary success (B-rationality) of each option is simply determined by their mean arithmetic fitness. Yet, if the risk is purely aggregate,[9] the principle of expected utility is prima facie no longer appropriate, as the expected number of offspring produced by the individuals is no longer the main determinant of their

---

[8] A paradigmatic example of aggregate risk is weather: indeed, if a population of individuals opt (say) for a wet-specialized strategy, then all of them will suffer an equal fitness loss in a dry year (and, equally, an equal fitness gain in a wet year). By contrast, a good example of an idiosyncratic risk is predation risk: indeed, in most cases, being caught by a predator does not imply that another individual with the same behaviour will also be caught by a predator.

[9] If the risk has both an idiosyncratic and an aggregate component, things are different, as we will see in the next section.

long-term evolutionary success. Instead, the *variance* in the number of offspring becomes the key determinant.

In Okasha's example, option B has a greater variance than option A (zero variance). Given aggregate risk, all of the individuals who have "opted" for this strategy enjoy or suffer *together* the same fitness consequences—good or bad. But, because of this aggregate component, all of the type B individuals will also end up representing a very small fraction of the global population. For even though bad runs are unlikely, they will definitely occur. Hence, even though A and B individuals have the same expected number of offspring, risk-averse (A) individuals are the ones that will ultimately dominate the population.

To account for this fact, a solution consists in using *geometric* mean fitness (instead of expected fitness) in order to measure the long-term evolutionary success of the different strategies (McNamara 1995). Thus, in Okasha's example, the geometric mean fitness G_A and G_B of strategies A and B are equal to:

$$G_A = 0.5 \times 5 \times 0.5 \times 5 = 6.25$$

$$G_B = 0.5 \times 9 \times 0.5 \times 1 = 2.25$$

One can see that $G_A > G_B$. So geometric mean fitness is clearly sensitive to variance. But unlike expected fitness, this measure no longer fits the arithmetic structure of the principle of expected utility—instead, natural selection, when maximizing geometric fitness, seems to follow a principle of *non*-expected utility. For this reason, one could be tempted to conclude that, at least in the event of purely aggregate risk, the B-rational option is actually *not* E-rational. But as we will now see, things are a bit more complicated.

5 B-rationality "violations" of E-rationality reconsidered

There are two possible ways of replying to the claim that B-rationality can (sometimes) *directly* induce violation of E-rationality. The first consists in denying that there is a violation of E-rationality in the case at hand; the second consists in denying the assumption that B-rationality is satisfied in the first place. Here, we illustrate both of these approaches, by respectively focusing on the two examples introduced in our previous section.

5.1 E-rationality is not violated: revisiting the case of McNamara et al. (2014)

In McNamara et al.'s paper, the authors conclude that E-rationality is violated as a result of B-rationality. A closer look at this example, however, shows that this apparent violation is but an artefact of a phenomenon known (in economics) as the "epistemic value of menus" (Sen 1993, 1997), a phenomenon which occurs when the composition of the choice set carries some information about the actual state of the world. Here, we use Sen's famous "tea-cocaine-home" example (Sen 1993, p. 502) to draw an analogy with McNamara et al.'s model.

Sen's example goes as follows. Suppose a distant acquaintance offers you to have a cup of tea at his home. You accept. In this case, the choice set is composed of two options, X and Y, where X is "having tea at a distant acquaintance's home" and Y is "stay home". Suppose further that this same acquaintance makes you the additional offer (say, later in the conversation) of having some cocaine at his home. This time, you decline, and decide to stay at home (you choose option Y). In this latter case, the choice set is now composed of X, Y and Z, where Z is the option "having cocaine at a distant acquaintance's home".

As one can see, the addition of a third option Z leads to an apparent "preference reversal" between options X and Y (in the first case, X is chosen, whereas in the second case, Y is chosen, so the axiom of independence of irrelevant alternatives appears to be violated). But, clearly, the fact of declining the offer *after having initially accepted it* does not constitute an instance of E-irrational behaviour. Rather, what happens is simply that the new composition of the choice set (the "menu") *informs* you about a relevant aspect of the world—namely that your acquaintance, contrary to what you formerly thought, is not a "decent" person, but an "undesirable" fellow. Hence, because your preferences ultimately depend on the actual state of the world, which, in turn, can be inferred from the composition of the menu, there is no relevant comparison between your preferences in the first menu (without cocaine) and your preferences in the second menu (with cocaine).

Bossert and Suzumura (2011), in a discussion of Sen's example, suggested to make a distinction between the *objects of choice* and the *objects of preference*, in order to reconcile the main standard of rational choice with the observed preferences of the agents. Applied to Sen's example, their distinction identifies three possible objects of choice, namely X, Y and Z, but four objects of preferences, namely:

(a) Having tea at a place presumed free of cocaine.
(b) Having tea at a place where cocaine is consumed.
(c) Having cocaine.
(d) Staying home.

Thus, when the subject is offered cocaine, what happens is simply that outcome (a) ceases to be a possible consequence of choosing X ("tea"), meaning that there is no violation of the axiom of independence of irrelevant alternatives in the first place; indeed, the preferences are here explicitly defined over (a), (b), (c) and (d), and not over the three options X, Y and Z.

A similar logic can be applied to McNamara et al.'s example. To illustrate this point, consider the pairwise choice between options A and C. In this setting, option A is chosen by the individual because, though C has the highest profitability for the first 10 units of time, no other option is available after that time (so A, overall, has the greatest value per unit of time). However, when option B is added to the choice set, the choice between A and C is reversed because, after choosing C, the individual can now choose B twice in a row and thereby maximize its rate of energy intake. At first sight, one could be tempted to conclude (like McNamara and his colleagues) that the independence of irrelevant alternatives is violated. But when the

objects of choice are distinguished from the objects of preferences (the fitness outcomes), it becomes clear that no violation of E-rationality is made: instead, what one observes is simply the individual *adjusting* its behaviour to the actual state of the world.

In this situation, more precisely, we have three objects of choice, A, B and C, but four objects of preferences (each defined over an interval of 20 units of time), namely:

(e)   Choosing C (10 units of time) then nothing else.
(f)   Choosing C (10 units of time) then B (5 units of time) and B again.
(g)   Choosing A (20 units of time).
(h)   Choosing B (5 units of time) then B three more times.

In the pairwise choice between A and C, outcome (f) is not a possible consequence of choosing C; for when faced with this choice, the individual "knows"—from the composition of the menu—that B won't be available later. However, when option B is added to the choice set, the uncertainty about the availability of B in the future is automatically "lifted": for outcome (e) ceases to be a possible (B-rational) consequence of choosing C, while outcome (f) becomes the only (B-rational) consequence of C. Hence, B-rationality no longer conflicts with E-rationality, as both are defined over the objects of preferences.

5.2 B-rationality is not satisfied in the first place: revisiting the case of risk-aversion

In the case of risk averse behaviours, a possible way of reconciling B-rationality with E-rationality could be to redefine the utility function associated with each outcome. But as stressed by Okasha, the "success" of this operation is, ultimately, conditional upon the composition of the risks.

When risk is aggregate, as we have seen, the criterion for B-rationality is given by geometric mean fitness, which is not, as such, compatible with the principle of expected utility and its arithmetic form. But this incompatibility, in the case of *purely* aggregate risk, can be easily circumvented by choosing to define biological utility as the logarithm of absolute fitness $\log(w)$. The reason is simple: the geometric mean of w is mathematically equivalent to the arithmetic mean of $\log(w)$. Hence, by choosing this latter measure instead of w, the arithmetic form constitutive of the expected utility principle is automatically recovered, and the existence of risk averse behaviours simply appears as a mere "consequence" of the concavity of the log function.

When the risks are mixed, however, things are more complicated, for the content of the log utility function is now *itself* an expectation over idiosyncratic risks. This can be seen by considering the following expression (Robson and Samuelson 2010), which gives the (correct) formula for the expected fitness ew associated with each option (though not "expected" in the sense of expected utility):

$$ew = \sum_{s} \log \left(\sum_{x} x \mathrm{p}(x|s) \right) f(s)$$

Here, f represents the probability distribution over aggregate risks; p represents the probability distribution over idiosyncratic risks; $f(s)$ measures the probability that the world is actually in state s (say, a harsh winter) and $\mathrm{p}(x|s)$ measures the probability that the individual get outcome x (say, a safe spot for foraging) given s. Now, in the RHS of this expression, the expectation within the log utility function corresponds to the mean absolute fitness $w(s)$ computed over the outcomes *in state s*. But, because this term is precisely encapsulated within the log function, it is impossible to extract the probabilities $\mathrm{p}(x|s)$ from it to determine the *marginal* probabilities associated with each possible fitness outcome. Consequently, the whole expression ew fails to satisfy a fundamental requisite of the principle of expected utility, which is the strict separability between the outcomes and their marginal probabilities (Machina 1989).

To overcome this problem, some authors (Grafen 1999; Curry 2001) have suggested that, by using *relative* fitness $r(s)$ as a measure of biological utility, a simple link between B-rationality and E-rationality could be recovered. But as pointed by Okasha (2011), this solution relies on *cognitively implausible* assumptions at the individual level. Indeed, unlike absolute fitness $w(s)$, relative fitness $r(s)$ is a function of the mean fitness of the population $\overline{w}(s)$ which, in turn, is a function of the global frequencies of the different behavioural types in the population. Using $r(s) = w(s)/\overline{w}(s)$ in place of the (cumbersome) log function of the previous expression, the criterion for B-rationality thus becomes:

$$ew = \sum_{s} \frac{w(s)}{\overline{w}(s)} f(s)$$

But the problem with this "maximand" is that its value *varies* as a function of the global frequency of individuals using the risky strategy; yet, as noted by Okasha, it is very unlikely that an organism could adjust its behaviour (more or less "risk-averse" or "risk-prone") to the global frequency of a given tendency within a population.[10] Hence, unless this latter possibility can be empirically established at the level of the individual organisms, we should assume that the hypothesis of B-rationality is *not* satisfied in the first place.

To this point, one could retort that, in the cases where frequency-dependent selection is at work, there is no need to claim that the individuals should be able to "adapt" their behaviour to the global frequencies at the evolutionary stable state. For, even with fixed behaviours, one should expect natural selection to favour the behavioural partition that corresponds to the (E-rational) mixed Nash equilibrium.[11] But in our view, this objection confuses the populational level with the individual

---

[10] More precisely, this is unlikely if we assume that the individuals are "playing against the field", and not against particular individuals, as in pairwise interactions (Maynard-Smith 1982). For in this latter case, the individuals could possibly use cues or signals correlated with the behaviour of their partner to adopt the best response.

[11] We thank an anonymous reviewer for having raised this important issue.

level, for the concept of rationality does not apply in the same way to natural selection and to individual organisms. Granted, if we characterize natural selection as the proper analog of a "rational agent", we should observe, at equilibrium, a set of behavioural frequencies corresponding (in the population) to the ratio of options given by the corresponding mixed Nash equilibrium, and admittedly, this would be so *whether or not* these behaviours are implemented in a plastic or in a rigid way by the individuals. But ultimately, this representation of natural selection as an E-rational agent turns out to be irrelevant to our problem as it does not help us to determine the conditions under which natural selection *realizes* each organism as an E-rational decision-maker, which is the problem we have been investigating in this paper. Thus, given that we are primarily interested in the E-rationality of individual organisms, the epistemic limitation mentioned above provides a good (empirical) reason for *not* assuming that natural selection will always favour individuals that behave like rational, i.e. adaptive, agents in frequency-dependent contexts.

6 Conclusion

The structural affinity between natural selection and rationality in economics encourages us to expect that organisms will respect the usual clauses of rational choice. To this extent, violations of these clauses in specific experiments by some animals appear surprising. We have characterized those violations as a clash between two forms of rationality, namely E-rationality and B-rationality, following Kacelnik (2006) terminology.

Apparent irrationalities in behavior (namely, discrepancy between being B-rational and E-irrational) can be understood under the perspective of selectionist explanations. Within this broad Darwinian framework, we distinguished two families of explanations. In the first one, E-rationality is seen, roughly, as a *by-product* of the selection for decision patterns that are, on average, adaptive in most selective environments met, (trade off hypothesis), or that were adaptive in past environments (mismatch hypothesis). This selection of irrational behaviours contrasts with the second family of explanations, where E-irrationality, it is argued, is not only related to what is selected, but moreover, is *itself* selected for its contribution to fitness.

In the last section, we challenged this latter kind of explanation. For the example provide by McNamara et al. we showed that the initial description of choices as violating transitivity or independence of irrelevant alternatives could be reformulated in favour of a description in which those choices remain E-rational. For the risk-aversion example, following Okasha (2011), we have seen that the assumption that choices are fitness maximizing and therefore B-rational in the first place can also be challenged, which makes the contradiction disappear since E-rationality is not expected from decision-makers that are not by nature maximizers. Therefore, even if in principle there could be selection for irrational decision making, in fact the cases presented as such are often controversial.

The last consequence of our study concerns the status of rationality itself. Emphasizing the nature of both natural selection and rational choice as maximizing

processes easily leads us to interpret rationality as one general kind of process, prevailing in both biology and economics. Under this perspective, while economics is the science of the optimal allocation of scarce resources, evolutionary biology, or at least behavioural ecology, can appear as the science of the optimal allocation of fitness, hence unifying both fields into a general science of rationality as optimality. Irrational behaviours are problematic in this view. Instead of this prospective unity of rationality diversified into economics and biology, it would suggest that the B-rationality and E-rationality are intrinsically different, and that the instances of "rationality" in these two concepts are not likely to be integrated into a general notion of rationality, but are almost homonymous.

However, analyzing irrational behaviours in more detail, and leaving aside cases where in fact B-rationality does not even occur, we see that those putative irrationalities can be conciliated with selection. Notwithstanding whether there are uncontroversial cases of selection *for* irrationality, as discussed in Sects. 4 and 5, the genuine cases where irrationality occurs can be seen as selection *of* irrationality (i.e. irrationality as a by-product), and, therefore, the prospect for maintaining a univocal concept of rationality across biology and economics remains intact.

E-rationality and B-rationality are indeed different in the sense that selection may favour outcomes (either as direct targets or as by-products) that depart from E-rationality—while being by definition B-rational. But this does not imply that the two concepts are wholly distinct. Actually, a consequence of our paper is that E-rationality and B-rationality can diverge in some specific cases, but that they both instantiate a specific kind of strategy-choice, allowing the modeler to predict outcomes. While selection is not a conscious choice maker, as Darwin emphasized when he wished to substitute "survival of the fittest" for "natural selection", recent behavioural ecology seems to have developed the theoretical and modeling consequences of the fact that selection and rationality are indeed both about choosing strategies.

References

Adams, F., & Garrison, R. (2013). The mark of the cognitive. *Minds and Machines, 23*(3), 339–352.

Akerlof, G. A., & Shiller, R. J. (2009). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton: Princeton University Press.

Bateson, M., Healy, S. D., & Hurly, T. A. (2002). Irrational choices in hummingbird foraging behaviour. *Animal Behaviour, 63*(3), 587–596.

Birch, J. (2015). Natural selection and the maximization of fitness. *Biological Reviews, 91*(3), 712–727.

Bossert, W., & Suzumura, K. (2011). Rationality, external norms, and the epistemic value of menus. *Social Choice and Welfare, 37*(4), 729–741.

Camerer, C. (2003). *Behavioural game theory: Experiments in strategic interaction*. New York: Princeton University Press.

Cosmides, L., & Tooby, J. (1997). Dissecting the computational architecture of social inference mechanisms. In G. Bock & G. Cardew (Eds.), *Characterizing human psychological adaptations (Ciba Symposium Volume #208)* (pp. 132–158). Chichester: Wiley.

Curry, P. A. (2001). Decision making under uncertainty and the evolution of interdependent preferences. *Journal of Economic Theory, 98*(2), 357–369.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An introduction to behavioural ecology*. Hoboken: Wiley-Blackwell.

Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. Cummins & C. Allen (Eds.), *The evolution of mind* (pp. 9–29). New York, NY: Oxford University Press.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences, 205*(1161), 581–598.

Grafen, A. (1999). Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. *Proceedings of the Royal Society of London B: Biological Sciences, 266*(1421), 799–803.

Hammerstein, P. (2002). Evolutionary adaptation and the economic concept of bounded rationality—a dialogue. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality. The adaptive toolbox* (pp. 73–82). Cambridge: MIT Press.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*(1), 81–91.

Haselton, M. G., & Buss, D. M. (2009). Error management theory and the evolution of misbeliefs. *Behavioral and Brain Sciences, 32*(6), 522–523.

Houston, A. I. (1997). Natural selection and context-dependent values. *Proceedings of the Royal Society B, 264*, 1539–1541.

Houston, A. I. (2012). Natural selection and rational decisions. In S. Okasha & K. Binmore (Eds.), *Evolution and rationality: Decisions, cooperation and strategic behaviour* (pp. 50–66). Cambridge: Cambridge University Press.

Houston, A. I., McNamara, J. M., & Steer, M. D. (2007). Do we expect natural selection to produce rational behaviour? *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 362*(1485), 1531–1543.

Hurly, T. A., & Oseen, M. D. (1999). Context-dependent, risk-sensitive foraging preferences in wild rufous hummingbirds. *Animal Behaviour, 58*(1), 59–66.

Kacelnik, A. (2006). Meanings of rationality. In S. L. Hurley & M. Nudds (Eds.), *Rational animals* (pp. 87–106). Oxford: Oxford University Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Kahneman, P. Slovic, & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Latty, T., & Beekman, M. (2011). Irrational decision-making in an amoeboid organism: Transitivity and context-dependent preferences. *Proceedings of the Royal Society of London B: Biological Sciences, 278*(1703), 307–312.

MacArthur, R. H., & Pianka, E. R. (1966). On optimal use of a patchy environment. *American Naturalist, 100*(916), 603–609.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature, 27*(4), 1622–1668.

Maynard-Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

McNamara, J. M. (1995). Implicit frequency dependence and kin selection in fluctuating environments. *Evolutionary Ecology, 9*(2), 185–203.

McNamara, J. M., Trimmer, P. C., & Houston, A. I. (2014). Natural selection can favour 'irrational' behaviour. *Biology Letters, 10*(1), 20130935.

Moran, P. A. P. (1964). On the non-existence of adaptive topographies. *Annals of Human Genetics, 27*(4), 383–393.

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector problem. *Evolution and Human Behavior, 26*, 88–105.

Okasha, S. (2011). Optimal choice in the face of risk: Decision theory meets evolution. *Philosophy of Science, 78*(1), 83–104.

Robson, A. J. (1996). The evolution of attitudes to risk: Lottery tickets and relative wealth. *Games and Economic Behavior, 14*(2), 190–207.

Robson, A. J., & Samuelson, L. (2010). The evolutionary foundations of preferences. In J. Benhabib, A. Bisin & M. Jackson (Eds.) *Handbook of social economics* (Vol. 1B, pp. 221–310). Amsterdam: North Holland.

Schuck-Paim, C. (2003). *The Starling as a rational decision-maker*. Oxford: University of Oxford.

Sen, A. (1993). Internal consistency of choice. *Econometrica: Journal of the Econometric Society, 61*(3), 495–521.

Sen, A. (1997). Maximization and the act of choice. *Econometrica: Journal of the Econometric Society, 65*(4), 745–779.

Shafir, S. (1994). Intransitivity of preferences in honey bees: Support for 'comparative' evaluation of foraging options. *Animal Behaviour, 48*(1), 55–67.

Stearns, S. C. (1992). *The evolution of life histories*. Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science, 3*(6), 358–361.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

West-Eberhard, M. J. (2005). Phenotypic accommodation: Adaptive innovation due to developmental plasticity. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 304*(6), 610–618.