Biology and the Theory of Rationality

1. Introduction

Philosophers since antiquity have been interested in the nature of rationality. A central concern in epistemology is to assess the rationality of our beliefs, while a central concern in practical philosophy is to assess the rationality of our actions. These topics are interesting partly because it is not clear what the relevant standards are for assessing the rationality of beliefs and actions. For example, it is often said that rational beliefs are ones which are "apportioned to the evidence", but what exactly does that mean? Does it imply that two individuals with the same evidence must be in identical credal states on pain of one of them being irrational? Similarly, it is often said that rational actions should reflect an agent's beliefs about how best to bring about the consequences they most desire; but what exactly does this mean? What if the agent does not know the likely consequences of the different courses of action open to her? What if the agent desires things that are harmful for her? Though we all have an intuitive grasp of what rational belief and action consist in, producing substantive analyses of these concepts has not proved easy.

Some progress on these issues comes from the theory of rational choice, the mainstay of modern economics. Rational choice theory offers a precisely defined, albeit rather "thin", notion of rationality. A rational agent's beliefs, on the standard picture, can be modelled by a subjective probability function over some set of alternatives ("states of the world"); when the agent gets new evidence, they update their probabilities by Bayesian conditionalization. As regards action, a rational agent chooses between alternative actions using expected utility maximization; i.e. by assigning utilities to the possible consequences of each action, and picking an action that maximizes expected utility with respect to their probabilistic beliefs. This picture of rationality involves a healthy dose of idealization, since real-life agents rarely have explicit probabilistic beliefs and almost never consciously compute expected utilities. However an ingenious argument, due originally to Ramsey (1931) and Savage (1954), shows that an agent whose binary choices satisfy certain fairly intuitive conditions necessarily behaves *as if* they had explicit probabilistic beliefs, an explicit utility function, and was aiming to maximize their expected utility.

Many philosophers define "rationality" in a richer sense than this, to mean that an agent has good *reasons* for their beliefs and actions, and that these reasons have been instrumental in causing the beliefs and actions. (Some would go further and require that a

rational agent be aware of these reasons.) Understood this way, rationality requires fairly sophisticated cognitive abilities, so is presumably the preserve of a few species, perhaps only *Homo sapiens*. By contrast, conforming to the consistency requirements of rational choice theory could in principle be achieved by an organism who lacked ‘‘”reasons” altogether but was capable of making behavioural choices. In a useful discussion, Kacelink (2006) refers to rationality in the sense of acting or believing on the basis of reasons as “PP-rationality” (standing for ‘‘philosophers and psychologists’’), and contrasts it with the ‘‘E-rationality’’ of economists, by which he means satisfying the standard principles of rational choice, such as expected utility maximization.

Can a biological perspective shed light on the nature of rationality? Scholars from a number of disciplines have suggested that it can. In philosophy, naturalistically-inclined thinkers at least since Quine (1969) have suggested that human rationality is the result of Darwinian selection; thus for example Dennett (1987) claims that “natural selection guarantees that most of an organism’s beliefs will be true, most of its strategies rational” (p.7). More recently, Sterelny (2003) argues that belief / desire psychology, which arguably underpins our capacity for rational thought and action, can be considered an adaptation to a ‘‘hostile environment’’ and has sketched an account of how it might have evolved; Godfrey-Smith (1996) argues similarly. In a different vein, authors such as Skyrms (1996) and Binmore (2005) have argued that evolutionary considerations can illuminate a variety of phenomena that traditional rational choice theory struggles to explain, such as the human sense of fairness and our capacity for altruism. A useful survey of philosophical work on the evolution / rationality connection is Danielson (2004).

In psychology, a number of authors have advocated a Darwinian approach to human cognition and decision-making, by focusing on the question of adaptive function. Thus Gigerenzer and colleagues argue that many aspects of human cognition which appear defective by traditional rationality criteria may generate adaptive behaviour in particular environments, so are thus “ecologically rational”. A recent paper in this vein by Hammerstein and Stevens (2014a), entitled ‘‘Six Reasons for Invoking Evolution in Decision Theory’’, argues that instead of the traditional axiomatic approach to rational decision, we should study decision-making using an evolutionary approach. They suggest that considerations about what is adaptive, rather than what is ‘‘rational’’ according to some idealized theory, will shed more light on how humans actually make decisions. A related argument is made by the evolutionary psychologists Cosmides and Tooby (1994), who argue that the mind comprises

evolved "modules" equipped for specific tasks, which enable "better than rational" behaviour. Useful surveys of this area include Gigerenzer and Selten (2001) and Hammerstein and Stevens (2014b).

In behavioural ecology, the branch of evolutionary biology which studies animal behaviour from a Darwinian basis, rationality concepts play an interesting role. Though this field focuses mainly on non-human animals, it has often borrowed models and concepts from rational choice theory and given them a biological twist. Typically this involves re-interpreting the utility function as a biological fitness function, and allowing natural selection rather than a rational agent to do the optimizing. Thus for example models of optimal foraging often assume that animals foraging for food behave like rational Bayesian agents, updating their ''beliefs'' on receipt of new information and choosing fitness-maximizing strategies (Houston and McNamara 1999). Similarly, Maynard Smith (1982) famously utilized concepts from classical game theory to shed light on social interactions among animals, giving rise to the field of biological game theory (see section 3 below). It is striking that rational choice models, which have often been criticized for assuming "superhuman" reasoning abilities, should prove so useful for understanding the behaviour of animals with only limited cognitive powers.

In cognitive and comparative psychology, there is considerable discussion of whether, and in what sense, the behaviour of non-human animals qualifies as rational. Researchers in this area often give intentional or ''belief-desire'' explanations of the behaviour of animals, including mammals and birds. For example, Clayton, Emery and Dickinson (2006) argue persuasively that the food caching and recovery behaviour of western scrub jays is most naturally explained by attributing to them beliefs and desires; that alternative non-intentional explanations fail, and that the jays' behaviour is therefore rational. Against this, it might be argued that the birds do not have beliefs in the full sense (perhaps because they lack language), or that even if they do, their behaviour is not rational since it is not reason-based in the requisite way. This issue turns in part on the correct interpretation of the empirical evidence and in part on the precise concept of rationality that is in play. A useful collection of papers in this area is Nudds and Hurley (2006); see also Andrews (2014) section 2.3.

In economics, there is a growing literature on the biological foundations of preferences. While most economic theorizing takes an agent's preferences (e.g. over consumption bundles) as a given, this literature asks what sort of preferences we should

expect to evolve by Darwinian selection. The underlying assumption is that human preferences stem from our evolved psychology, so should admit of a Darwinian explanation. Thus for example Robson (1996) studies the evolution of attitudes to risk, producing the striking finding that in certain circumstances, agents whose preferences violate the axioms of expected utility theory should enjoy a selective advantage (see section 5). In principle, this type of argument could help explain why the actual behaviour of humans seems to systematically depart from the predictions of rational choice theory. A useful overview of work in this field is Robson and Samuleson (2011).

A proper survey of the diverse lines of investigation described above would be beyond the scope of a single paper (and most probably, author). Here my focus is on overarching philosophical and conceptual issues. In section 2, I examine the idea that biology supplies an alternative evaluative yardstick for assessing beliefs and actions, distinct from the yardstick employed in traditional discussions of rationality. In section 3, I look briefly at the concept of ecological rationality and its implications for the study of the human mind. In section 4, I examine the link between evolution and rational choice, focusing on the idea that Darwinian fitness can supply some "meat" to the abstract utility function of rational choice theory. In section 5, I examine the idea that evolution and rationality can "part ways", i.e. that evolutionarily successful behaviour may fail to coincide with rational behaviour. Section 6 concludes.

2. Biology and the ''Yardstick'' of Rationality

Rationality is a normative notion. Rational beliefs and actions are ones which conform to the norms of belief formation, belief change, and choice of action, whatever exactly they are. (This is so whether we are talking about ''PP-rationality'' or ''E-rationality'' in Kacelnik's terms). Thus to call a belief or action rational is not simply to describe it but also to evaluate it. Indicative of this normativity is the fact that it makes sense to ask what beliefs a person *should* have, given their evidence, and what action they *should* choose, given their aims (or perhaps, given the aims that we think they should have.) The source of this normativity is a deep philosophical issue, according to some authors; but fortunately we can leave this matter aside. For the moment, the point is simply that inherent in the idea of rationality is the idea that an agent should believe or act in a certain way, and thus the possibility that the agent's actual belief or action will fail to be as it should, hence irrational.

One way to see the relevance of biology to rationality theory is to note that evolutionary biology suggests its own normative standard by which to assess actions (and indirectly, beliefs). Consider a male organism in a sexual species who is trying to attract a mate. A variety of possible mating strategies exist, e.g. performing a showy display, engaging in male-male combat, or trying to take control of another male's harem, each of which will have different consequences for the organism's reproductive success (or ''fitness''). This suggests a natural way of normatively evaluating the organism's choice of strategy. As well as asking which mating strategy the organism *does* actually adopt, we can also ask which strategy it *should* adopt, i.e. which strategy will be fitness-maximizing in the relevant environment, or evolutionarily optimal. If the organism chooses a sub-optimal strategy, it makes sense to say that the organism has failed to do what it should have done, or has failed to achieve the ''goal'' of maximizing its reproductive success.

The fact that evolutionary biology supplies its own yardstick of normative evaluation, based on the calculus of Darwinian fitness, yields a notion of ''biological rationality'' (cf. Kacelnik 2006), that is logically distinct from the rationality notions used in other disciplines, but may nonetheless bear interesting relations to them. Since biological rationality is all about enhancing one's fitness, the notion applies in the first instance to behaviours, or choices. As such the notion is applicable to any organism capable of behavioural plasticity. A bacterium that swims towards a chemical gradient has made a ''choice'' about which direction to swim in, and it makes sense to ask whether its choice is the ''correct'', i.e. fitness-maximizing, one. The notion can also be applied to beliefs and desires, so long as the organisms in question are capable of having them, for these mental states give rise to behaviour. Thus in principle, various aspects of human cognition and decision-making can be evaluated by the yardstick of biological rationality (see section 3).

Biological rationality appears logically independent, in both directions, of rationality in the sense of having reasons for one's beliefs and actions. An agent's beliefs and actions might be suitably reason-based and yet not enhance their biological fitness; conversely, an agent's beliefs and actions might be fitness-enhancing and yet not based on good reasons, perhaps because they lack the capacity to have reasons at all. What about rationality in the sense of conformity to the norms of rational choice theory? It seems obvious that this need not imply biological rationality: an agent may have consistent preferences that are detrimental to their biological fitness, as many modern humans arguably do. But the converse inference, from biological rationality to conformity to rational choice norms, has often been defended

(eg. Gintis 2009 p.7, Kacelnik 2006, Chater 2012). This inference seems reasonable: if an organism displays adaptive behaviour, so chooses actions that maximizes its fitness, then presumably it is behaving like a utility-maximizing agent whose utility function is simply its fitness function? In fact matters are not quite so simple, for reasons discussed in sections 4 and 5.

Some philosophers might dispute whether biological rationality counts as a genuine species of rationality, on the grounds that it is really just another name for adaptiveness, or fitness-maximization. According to this objection, the sense of ‘‘should’’ in which an animal should perform a biologically rational action carries no real normative force, and is not interestingly similar to the sense of ‘‘should’’ in which humans should base their beliefs on the evidence, or conform to the dictates of rational choice theory, for example. After all, wherever the notion of adaptive function applies then it makes sense to talk about malfunctioning, or not doing the ‘‘correct’’ thing, as proponents of teleosemantics have long stressed; but malfunctioning is not usefully equated with irrationality. So biological rationality does not deserve its name, the objection goes.

In response, it must be granted that the notion of adaptive function applies in contexts where talk of rationality or irrationality would be inappropriate. If an organism’s digestive system malfunctions, for example, it makes good sense to say that the system has not done what it should do, but this is not a *rational* shortcoming. The operation of the digestive system is too automatic for such a characterization to be useful. However matters are different when we are dealing with behaviours or actions, particularly if the organism in question displays considerable behavioural plasticity, or is capable of learning about its environment and modifying its behaviour to suit the circumstances, as many birds and mammals can. Animal behaviour of this sort is objectively similar (at a suitable grain of description) to human behaviour, and in some cases is homologous with it, despite the desire of some philosophers to see a chasm between humans and non-humans. Where such behaviour is concerned, the normativity that derives from the notion of adaptive function is plausibly regarded as a type of rationality, or proto-rationality.

This point can be bolstered by recalling two facets of the traditional rationality concept discussed by philosophers. Firstly, rational action is *goal-directed* action, in which an agent is trying to achieve an end. An action qualifies as rational to the extent that it serves the agent’s end (or is believed by the agent to do so.) Much animal behaviour appears

unambiguously goal-directed – think of a bird collecting sticks in order to build a nest, or a primate sharpening a tool in order to crack a nut, or a honey bee performing a waggle-dance in order to communicate the location of a nectar source. It is difficult to make sense of such behaviour without the assumption that it is goal-directed (certainly in as ‘‘as if’’ sense, and arguably in a stronger sense). In recognition of this, behavioural ecologists frequently use an intentional idiom (e.g. ‘‘wants’’, ‘‘tries’’, ‘‘knows’’, ‘‘communicates’’) to describe and explain animal behaviour; this idiom is typically regarded as neither metaphorical nor dispensable. Calls to banish the intentional idiom from the study of animal behaviour (e.g. Kennedy 1992) have been noticeably unsuccessful. From this perspective, the evaluation of behaviour in terms of its biological rationality looks like a *bona fide* species of rational evaluation.

Secondly, as McDowell (1994) argues, following Davidson (1984), when we give a folk psychological explanation of an agent's action or belief, our explanation makes the belief or action intelligible by rationalizing it; this is quite different from an explanation in physics, in which in which a phenomenon is made intelligible by showing that it had to happen as a matter of natural law. I suggest that this lends support to the view that biological rationality is a genuine type of rationality.[1] For when we explain an organism's behaviour in terms of its biological rationality, e.g. a chimpanzee fashioning a tool from a twig in order to catch termites, this yields exactly the sort of intelligibility that McDowell treats as definitive of rationalizing explanations. We are able to see how the behaviour makes sense, or is appropriate, in terms of the organism's goal (acquiring food), which itself subserves the ultimate goal of maximizing fitness. The type of understanding that we get of the organism's behaviour is more akin to the type of understanding we get from intentional explanation than from physical explanation.

One distinctive feature of the biological rationality concept is that it is externalist. A behaviour counts as biologically rational if it is fitness-maximizing or adaptive, which depends on the environment. Craving high calorie foods was adaptive in the Pleistocene environment of our hominid ancestors, but for humans in today's environment is not. By contrast, rationality in the sense of having reasons for one's beliefs and actions, or in the sense of conforming to the consistency conditions of rational choice theory, are internalist matters. Whether an agent is rational in either of these senses depends on how things are ‘‘in

[1] This is somewhat ironic, given that both McDowell and Davidson treat rationality as the preserve of human beings.

their head'', not in the external environment; so a suitably intelligent agent should be able to achieve rationality simply by a process of self-reflection and amelioration. For biological rationality, by contrast, the world must cooperate too.

Though biological rationality may be logically independent of rationality in the other senses discussed above, it is tempting to think that empirically, it must be related somehow to them. Creatures who act and believe for good reasons, or whose choices conform to rational choice norms, will generally enjoy a selective advantage over ones that do not, the suggestion goes; thus rationality in these senses, and the cognitive equipment necessary for them, are themselves Darwinian adaptations. Therefore, for the most part, beliefs and actions that are rational in the philosophical or economic senses are also likely to be biologically rational – or else natural selection would never have led to them. Dennett (1987) gives voice to this sentiment in the quotation above, when he asserts that natural selection ensures that most of our beliefs will be true and most of our strategies rational (cf. Stephens 2001).

This conjecture may be correct, but it is an empirical issue and potential counterexamples abound. One interesting counterexample comes from D.S. Wilson's work on the evolution of religion. Wilson (2002) opposes the modern liberal idea that religious belief is simply a rational pathology, or the result of our usually accurate belief-forming processes going awry. Instead of assessing religious beliefs against the yardstick of factual truth or epistemic rationality – by which they inevitably fall short – he argues that we should instead use an adaptationist yardstick. Wilson claims that religious believers are motivated to engage in pro-social actions to fellow group members, resulting in group-level benefits. Thus a process of between-group selection would have favoured religious over non-religious groups, he argues. If Wilson's (controversial) theory is true, then it renders religious beliefs and practices intelligible by showing that they ''make sense'' when judged by the criterion of fitness maximization, despite violating the usual norms of rational belief formation.

To summarize so far: evolutionary biology suggests a way of normatively evaluating actions and beliefs, by how well they promote an organism's fitness in its environment, that is distinct from the type of normative evaluation traditionally invoked in philosophy and in rational choice theory. Though *sui generis*, biological rationality is still a *bona fide* type of rationality, since it enables us to "make sense" of the beliefs and actions of both humans and non-humans by showing how they help to fulfil their evolutionary goal.

3. Humans and Ecological Rationality

Proponents of "ecological rationality", notably Gerd Gigerenzer, Peter Todd and colleagues, focus primarily on human psychology and cognition (Gigerenzer 2010, Todd, Gigerenzer et. al. 2012). Their theory incorporates aspects of biological rationality, in that it emphasizes successful performance in particular environments, but has a distinct focus. Gigerenzer's point of departure is Herbert Simon's concept of "bounded rationality", which stresses that humans do not have unlimited computational abilities, so cannot implement sophisticated optimization algorithms. Thus we rely on heuristics, or rules-of-thumb, to make decisions and solve problems. These heuristics are special-purpose and are tailored to specific environments, allowing them to exploit environmental regularities. (For example the "recognition heuristic" says that if choosing between two objects, one familiar the other unfamiliar, choose the familiar one. In an environment full of dangerous objects, this heuristic makes sense.) These "fast and frugal" heuristics are computationally cheap but get the job done.

Ecological rationality theorists emphasize the domain-specific nature of the heuristics which guide human decision-making. A heuristic helps us with a particular task, e.g. determining whether a social partner is honest. Different tasks call for different heuristics, so the human mind is an "adaptive toolbox", Gigerenzer and Selten (2001) argue. By contrast, traditional rational choice theory is a domain-general approach: the maximize expected utility rule can be applied to any choice problem, and the rules of probability can guide uncertain reasoning about any subject matter. This emphasis on domain-specificity is also a theme in the work of evolutionary psychologists Cosmides and Tooby; they argue that on general Darwinian grounds we should expect the mind to be composed of specialized modules, as this allows more efficient problem-solving than applying an all-purpose "general intelligence" (e.g. Cosmides and Tooby 1994). This inference – from a Darwinian premise to a conclusion about the structure of the mind – seems plausible, but ultimately the issue must be settled by direct psychological and neurobiological evidence.

Ecological rationality theorists paint an optimistic picture of human psychology. This contrasts with the emphasis on "cognitive biases" by theorists such as Kahneman and Tversky, who document systematic departures from the norms of rational choice and probability theory (Kahneman 2011, Kahneman and Tversky 2000, Kahneman, Slovic and Tversky 1982). According to these theorists, humans commit basic probabilistic errors,

exhibit time-inconsistency in their inter-temporal choices, commit the base-rate fallacy, display "loss aversion", "uncertainty aversion", and are prone to an alarming variety of "framing effects". These results, which are experimentally well-confirmed, are often interpreted as showing that humans are "just not irrational". From a biological perspective this is somewhat puzzling, as it is hard to see why evolution would favour creatures prone to such biases. However ecological rationality theorists offer a different picture. Relying on simple heuristics, rather than attempting to implement optimization, is an efficient way of solving problems. In our evolutionary past there was a premium on making quick decisions and choices; so using a simple heuristic (rather than searching through all the options looking for the ''best'', for example) was an adaptive strategy given our limited computational powers. In their natural settings such heuristics work well, but applied out of context they can make us look irrational.

To the extent that this line of argument is successful, it makes it more intelligible, in broad biological terms, why human reasoning and decision-making exhibit some of the features that they do. However, this is different from showing that the *specific* violations of rational choice precepts found by Kahneman, Tversky and others were to be expected. It is one thing to be able to explain, as Gigerenzer and colleagues arguably can, why humans do not make choices by explicitly trying to compute expected utilities, relying instead on simple shortcuts; but this does not explain the specific violations of expected utility maximization that have been found, such as displaying the Ellsberg preferences in choice under uncertainty, or using hyperbolic discounting in inter-temporal choice, for example. It is conceivable that these and related phenomena could be accounted for in terms of ecological rationality, but to date they have not been.

Proponents of ecological rationality are often rather disparaging of probability theory and rational choice theory. They regard the latter as *a priori* philosophical and mathematical exercises which do not help the quest to understand real-life decision making and cognition. Gigerenzer and colleagues argue that an agent who relies on ecologically rational heuristics for making choices will often outperform an agent who tries to conform to the decision-theoretic ideal, at least in the particular environments for which the heuristics were tailored. So not only is rational choice theory unhelpful for scientists seeking to understand human psychology, it is also unhelpful for agents themselves. Cosmides and Tooby (1994) argue similarly.

At times ecological rationality theorists go further, and argue that probability theory and rational choice theory are incorrect even as normative ideals, not merely that they are poor descriptions of how actual human cognition works. Thus Gigerenzer and Selten (2001) say that their theory "provides an alternative to current norms, not an account that accepts current norms and studies when humans deviate from these norms…bounded rationality means rethinking the norms as well as studying the actual behaviour of minds and institutions" (p.6). The suggestion, in short, is that the norms of traditional rational choice theory constitute an inappropriate standard by which to judge creatures which have evolved to be ecologically rational.

This negative attitude towards rational choice theory is by no means mandatory for those persuaded that adaptive considerations can illuminate the study of rationality. Indeed it is perfectly possible to hold, with the philosophical mainstream, that deductive logic and probability theory yield correct norms of rational belief, and decision theory correct norms of rational action, while at the same time holding that biological or ecological rationality constitutes a different standard by which our beliefs and actions can be normatively evaluated. I suggest that this attitude – permitting a plurality of valid rationality concepts – is more reasonable. We should allow that rationality in the sense of having good reasons for one's beliefs and actions, and rationality in the sense of conformity to rational choice precepts, are both valid forms of normative assessment; while also allowing that our beliefs and actions can be assessed in terms of ecological / biological rationality.

I suspect that the hostility of some ecological rationality theorists towards rational choice theory stems from the tendency, particularly among economists, to use the assumption of ideal rationality to build what are meant to be descriptively accurate models of human behaviour. This is certainly a questionable way to proceed, given that experimental work shows clearly that humans systematically violate the rational choice norms in at least some contexts (e.g. Ariely 2008). Given this fact, the idea of basing a science of human decision-making on Darwinian principles, rather than on the abstract axioms of decision theory, is undeniably attractive. However it does not follow that we should jettison decision theory and probability theory as normative ideals; but only that we should not assume without evidence that they are descriptively valid.

4. Utility and Fitness

Rational choice theory is sometimes criticized for relying on a purely abstract utility concept. To say that rational agents maximize their utility is not to say much, the criticism goes, since ‘‘utility’’ is in effect defined as whatever an agent wants. One version of this criticism goes further and alleges that utility-maximization is both empirically empty and normatively silent, since virtually anything that an agent does can be reconciled with it. This criticism is arguably overstated, particularly for choice under uncertainty, since the axiomatic conditions that an agent’s choice behaviour must obey for them to be describable as an expected utility maximizer are not trivial; but it is nonetheless true that the doctrine of utility-maximization provides little insight into why agents act as they do, or the reasons behind their choices. This is partly why traditional philosophical work on “practical reason” makes little use of utility theory.

If we are persuaded by the idea of a Darwinian approach to rationality, then a natural hope is that Darwinian fitness may put some “meat” on the abstract utility function of the rational choice theorists. To see why, consider a typical case of goal-directed animal behaviour: a foraging bird moving from one food patch to another as its rate of food intake declines. Moving patch incurs significant costs and risks, but may still be the best thing to do if food becomes too scarce in the current patch. So the bird needs to settle on a strategy for when to move from one patch to another. Suppose that the bird’s foraging behaviour has been honed by natural selection and so is biologically rational: it implements the strategy that will maximize its expected reproductive success, given the information it has. Armed with this knowledge, a scientist-observer can make precise sense of the bird’s behaviour, which might otherwise appear inexplicable or random.

The key point is this. The bird’s behaviour becomes explicable once we posit a *specific* goal, namely maximizing reproductive success (or some proxy for it); we know from evolutionary theory that behaviour directed towards this goal is a likely (though not inevitable) outcome of Darwinian selection. Thus our foraging bird is behaving like a utility-maximizer of the sort described by rational choice theory, *but whose utility function is of a very specific sort*. The bird behaves ‘‘as if’’ it cares about maximizing its expected reproductive success. Merely hypothesizing that the bird’s behaviour maximizes expected utility *modulo* some utility function or other, so satisfies the canons of rational choice, on its own explains rather little. It is the additional hypothesis that the bird’s utility function is its

fitness function that enables us to explain and predict its behaviour. This is the sense in which a biological perspective can put flesh on the bones of the utility function.

This observation tallies with the way that game-theoretical models, in particular, have been deployed in a biological context. Consider a simple two-player simultaneous game depicted in Figure 1, in which each player has two (pure) strategies at their disposal. The entries in each cell denote the payoffs to (player 1, player 2). In traditional game theory, these payoffs are assumed to be utilities; the assumption is that each player wants to maximize their (expected) utility. The game below has two pure-strategy Nash equilibria: (Top, Left) and (Bottom, Right), yielding payoffs of (2,2) and (1,1) respectively. Classical game theory offers these as the ‘‘solutions’’ of the game, and predicts that one of them will be observed; at such an equilibrium, each player is choosing the strategy that maximizes their payoff conditional on their opponent's strategy, so has no unilateral incentive to deviate.

	Left	**Right**
Top	(2, 2)	(0, 0)
Bottom	(0, 0)	(1, 1)

Figure 1: A game with two pure-strategy Nash equilibria

Beginning with Maynard Smith (1974, 1982), biologists have taken models of this sort and given them a biological twist, by interpreting the payoffs as fitnesses rather than utilities. So interpreted, the model describes a social interaction between two organisms, the outcome of which augments each organism's fitness by the relevant amount; so an organism's overall fitness depends both on its own strategy and that of its social partner. On the simplest assumption, each organism's strategy is genetically hard-wired and faithfully transmitted to its offspring. (Alternatively, the organisms may exhibit behavioural plasticity and be capable of choosing a strategy depending on an environmental cue.) Biologists typically imagine a large population of organisms, evolving by natural selection, and ask which strategy will come to dominate the population. Under reasonable assumptions, it can be shown that the population will usually reach an evolutionary equilibrium, corresponding to a Nash equilibrium of the original game.[2] Unlike in classical game theory, where an equilibrium is meant to result from a process of rational deliberation by intelligent agents, in

[2] See for example Weibull (1995) for a careful account of these assumptions.

biological game theory an equilibrium is reached as a result of a dynamical process, namely the differential proliferation of the fittest strategies.

This illustrates the fact that utility and fitness play isomorphic roles, in rational and biological game theory respectively. The former is the quantity that determines which strategy a rational agent will choose; the latter is the quantity that determines which strategy Darwinian evolution will program organisms to choose. When rational agents choose utility-maximizing strategies this leads to an equilibrium in rational deliberation; when organisms choose fitness-maximizing strategies this leads to an equilibrium of an evolutionary process. This consideration, and more generally the close analogy between the fitness-maximizing paradigm of evolutionary biology and the utility-maximizing paradigm of economics, has led many authors to see a deep connection between evolution and rationality theory (cf. Maynard Smith 1974, Stearns 2000, Grafen 2006a, Orr 2007, Okasha 2011).

The suggestion that utility and fitness are isomorphic in this way is appealing, but needs qualifying for three reasons. Firstly, it is not always clear what the analogue of the rational agent actually is, in a biological context. Usually it is individual organisms that engage in goal-directed behaviour, and whose choices may thus be evaluated in terms of biological rationality; but in other cases it is groups of organisms (or ‘‘superorganisms’’) that are the locus of goal-directed action, e.g. the co-ordinated behaviours of certain social insect colonies (cf. Seeley 1996, 2010.) In other cases still, involving conflicts of interest between the genes within an organism, the entity that has a ‘‘strategy’’ and is thus akin to a rational agent is the gene itself (cf. Haig 2012).[3] The question of which biological unit should be treated as agent-like (and why) is closely related to the discussion of ‘‘levels of selection’’ in evolutionary biology (Okasha 2006, Gardner and Grafen 2009).

Secondly, utility and fitness are measurable on different scale-types. In rational choice theory, utility is generally taken to be either ordinal or cardinal, depending on the problem at hand; in biology, fitness is generally treated as a ratio-scaled quantity, for the zero point of fitness is meaningful; so it makes sense to say that one strategy (or genotype) is twice as fit as another (cf. Grafen 2007). One *might* think there is a further disanalogy in that utility is usually taken not to be interpersonally comparable, while the whole point of the fitness concept is to compare the fitness of different individuals. But on most natural way of

3 This is known as ‘‘intra-genomic conflict’’, and arises because the genes in a sexually reproducing organism are not transmitted en masse to their offspring.

formulating the utility / fitness connection, there is a single fitness function for all individuals in the population, mapping strategies (or profiles of strategies, in the game-theoretic case) onto fitness. Different organisms play different strategies, hence receive different fitness payoffs; but this is simply the analogue of a rational agent receiving a different utility payoff from different outcomes, which involves only *intra*personal comparison.

Thirdly and most importantly, the appropriate definition of "fitness" is a subtle issue in biology, and depends on modelling assumptions. In the simplest evolutionary scenarios, expected lifetime reproductive success is the right fitness measure; natural selection favours organisms whose behaviour maximizes this quantity. Many phenotypic traits can be understood in terms of their contribution to maximizing expected reproductive success. However in more complicated scenarios matters are different. For example, if organisms engage in social interactions, then it is necessary to take account of the effect of an organism's actions on its genetic relatives, so Hamilton's ''inclusive fitness'' becomes the relevant measure (cf. Hamilton 1964, Grafen 2006a). If there is class structure in a population, e.g. individuals belong to different age-cohorts, then the appropriate fitness measure is different again, for it is necessary to weight offspring by their ''reproductive value'' (Charlesworth 1994, Grafen 2006b). So we cannot assume *a priori* that we know which quantity (if any) evolved organisms will behave as if they are trying to maximize (cf. Mylius and Diekmann 1995).

This consideration complicates the fitness / utility analogy but does not invalidate it altogether. For the basic Darwinian idea that natural selection will often give rise to adaptive behaviour is a mainstay of evolutionary biology, and enjoys broad empirical support. Many organismic traits, including behaviours, are manifestly there because they enhance the organism's "fit" to its environment. The fact that the appropriate quantitative measure of "fit" depends on the details of our evolutionary model shows that natural selection is a more complicated process than was once thought, but does not undermine the idea that adaptation to the environment, that results from selection, is a pervasive feature of the living world. To the extent that such adaptation occurs, it is legitimate to regard adapted organisms as akin to utility-maximizing rational agents trying to maximize their fitness, with the caveat that "fitness" must be defined appropriately for this idea to work and that different definitions may be needed in different cases.

4. Can Evolution and Rationality "Part Ways"?

We noted above that philosophers such as Quine and Dennett have argued that rational beliefs and behaviour are the likely outcome of natural selection. However against this, a number of authors have argued that considerations of rationality may sometimes "part ways" from considerations of fitness-maximization, to use an expression from Skyrms (1996). This is a striking suggestion, raising the prospect of an evolutionary explanation for why humans sometimes depart from traditional canons of rationality.

Skyrms illustrates this "parting of ways" with a simple Prisoner's Dilemma game, as in Figure 2. In a rational choice setting, in which the payoffs denote utilities, it is widely agreed that in the one-shot game the rational agent should play D (defect), as it strongly dominates C (cooperate). Thus the expected utility of playing D must exceed that of C. This is so even if the agent believes that its opponent is likely to play the same strategy as itself, presuming the truth of ''causal decision theory'' (Lewis 1981), as the two players are causally isolated.

		Player 2	
		C	**D**
Player 1	**C**	**(6, 6)**	**(0, 10)**
	D	**(10, 0)**	**(2, 2)**

Figure 2: Prisoner's Dilemma

Suppose we now transpose to an evolutionary setting and consider a large population of organisms engaged in a one-shot pair-wise interaction; the payoffs now represent increments of (personal) fitness. Which type has the higher fitness? As Skyrms observes, this depends on the pairing assumption that we make. Under random pairing, in which the probability of having a C partner is same for both types, it is obvious that type D must be fitter. The expressions for the fitnesses of the two types are then:

$$W_C = 6.P(C) + 0.P(D)$$

$$W_D = 10.P(C) + 2.P(D)$$

where P(C) and P(D) denote the probabilities of being paired with a co-operator and a defector respectively; these probabilities are given by the overall frequency of each type in the population. As Skyrms notes, these expressions for expected fitness are identical to the corresponding expressions for the expected utility in the rational choice context, calculated using standard (Savage-style) decision theory. Under random pairing, the type with the highest expected fitness chooses the action that confers the highest expected utility, so evolutionarily optimal behaviour is identical to rational behaviour.

Skyrms observes that matters are different if there is correlated pairing. We must then calculate the expected fitness of each type using the conditional probabilities of having a partner of a given type, which may differ for co-operators and defectors. The resulting expressions are:

$$W_C = 6.P(C/C) + 0.P(D/C)$$

$$W_D = 10.P(C/D) + 2.P(D/D)$$

where P(X/Y) denotes the probability of having a partner of type X, given that one is of type Y oneself. It is easy to see that if the correlation is strong enough, i.e. the conditional probability of having a C partner is sufficiently greater for C types than D types, then the C type may be fitter overall, and so spread by natural selection.[4] Skyrms concludes that with correlated pairing, "rational choice theory completely parts ways with evolutionary theory. Strategies that are ruled out by every theory of rational choice can flourish under favourable conditions of correlation" (1996 p. 106).

Sober (1998) develops the same point slightly differently, in the context of discussing what he calls the "heuristic of personification" in evolutionary biology. This heuristic is the idea that "if natural selection controls which of traits $T, A_1, \ldots, A_n$ evolves in a given population, then T will evolve, rather than the alternatives, if and only if a rational agent who wanted to maximize fitness would choose T over $A_1, \ldots, A_n$" (p. 409). Sober maintains that this heuristic is usually unproblematic but fails in certain contexts, one of which is the one-shot Prisoner's dilemma. The rational agent will never play co-operate, since it is strictly dominated, Sober reasons; however it is possible that natural selection will favour co-operate

[4] This is an instance of the statistical phenomenon known as ''Simpson's paradox''.

over defect if the requisite correlation exists. Thus the heuristic of personification fails: the rational strategy and the evolutionarily optimal strategy do not coincide.

These arguments are intriguing, but there is an obvious response, developed in detail by J. Martens (forthcoming). In the Skyrms / Sober model, there is no particular reason to equate the rational agent's utility function with its personal fitness function. Indeed evolutionary biology teaches us that in social settings, the relevant fitness measure is not personal fitness but *inclusive* fitness, as noted above. To calculate an organism's inclusive fitness, we need to take account of the effect of the organism's action on other members of the population, weighted by the ''coefficient of relatedness'' (denoted ''*r*'') between them. This coefficient is a measure of the genetic (and thus strategic) correlation between them; in the current context, the natural measure of ''*r*'' is [*P*(C/C) – *P*(D/C)].[5] It is straightforward to show that if a rational agent's utility function depends suitably on their inclusive fitness, then the Skyrms / Sober "parting of ways" disappears.

This particular "parting of ways" argument therefore does not succeed. Skyrms and Sober's model does not show that irrationality will evolve but rather that ''other regarding'' preferences will evolve – organisms will appear to care about the biological fitness of others as well as themselves. However there are other suggestions in the literature for how irrational behaviour may evolve. For example, Robson (1996), in an intriguing analysis, shows that organisms whose choice behaviour violates the axioms of expected utility theory will often enjoy a selective advantage, so will evolve in a population. This remarkable result arises from the existence of "aggregate risk", which refers to risks that are correlated across members of a biological population, e.g. bad weather. From a rational choice perspective, it should not make any difference to an agent whether a given risk is aggregate or not; but from an evolutionary perspective it does, given that what matters in evolution is reproductive success relative to the rest of the population. This is why Robson's model appears to yield the evolution of irrationality.

As with the Skyrms / Sober argument, however, it has proven possible to restore the connection between evolution and rationality in Robson's model by judicious choice of utility function (though the necessary ''fix'' in this case is far from obvious). Grafen (1999) and Curry (2001) both show that if an organism's utility, in each state of nature, is defined as its

[5] This is a special case of one standard definition of ''*r*'' in evolutionary theory, namely the linear regression of recipient genotype on actor genotype.

fitness divided by the average population fitness in that state, i.e. its relative fitness, then evolution will in fact favour maximization of expected utility after all; since expected relative fitness is the appropriate criterion of evolutionary success in the presence of aggregate risk. The key point is that, with aggregate risk, behaviour which fails to maximize an organism's expected absolute fitness may nonetheless maximize its expected relative fitness. So in theory, Robson's "parting of ways" can also be eliminated; though empirically, the idea that evolution could program an organism to care about its relative fitness is questionable, given that relative fitness depends on the actions of others so is not within an individual organism's control (cf. Okasha 2011).

It is tempting to suggest that the moral of the two cases above generalizes, i.e. that *any* putative "parting of ways" between evolution and rationality can in principle be avoided by suitable choice of utility function. However, there is no theoretical reason to think that this must be true. Moreover a number of authors have successfully developed models in which clearly irrational behaviours, for example intransitive choices, are favoured by natural selection and in which there is no obvious way to "restore" rationality by suitable choice of utility function (Houston, McNamara and Steer 2007). Thus it would be premature to conclude that a "parting of ways" argument cannot succeed, even given the latitude of defining an agent's utility function as we please. This issue needs to be judged on a case-by-case basis.

The "parting of ways" idea discussed above should be sharply distinguished from the quite different idea that humans derive positive utility from things that do not enhance their biological fitness (personal or inclusive). Empirically this clearly seems to be so: modern humans often have preferences for things that are neutral or detrimental to their fitness, e.g. sky-diving, contraception, or reading philosophy books. This is an interesting phenomenon, however it need not involve any irrationality in the sense of a violation of rational choice norms, so does not involve any "parting of ways" in the above sense. I conclude by briefly discussing the phenomenon.

From a biological perspective, is it possible to explain why humans derive utility from things that are detrimental to their fitness? Opinions on this issue differ. One response is that that human preferences are heavily dependent on learning and culture, exhibiting extensive cross-cultural variation; thus preferences are not under tight genetic control so are not susceptible to biological explanation. This may be partly correct, but it pushes the question

one step further back. Why did evolution make humans susceptible to acquiring preferences, by learning or cultural transmission, which would cause them to behave in ways that harm their biological fitness? Was it an unintended side effect of selection for the ability to learn, for example?

One interesting take on this issue comes from Sterelny (2012), who argues that at a certain point in hominin evolution, we changed from being "fitness maximizers" who desired things that are good for our genes, to being "utility maximizers" who desired things that are non or even maladaptive. Sterelny attributes this change to the shift from small-scale to mass society. In traditional small-scale societies, cultural transmission is primarily vertical, from parents to offspring, but as societies got larger, horizontal transmission became dominant. So individuals became susceptible to acquiring maladaptive beliefs and preferences by horizontal means. Moreover, in mass society the power of cultural group selection declines, so the filtering mechanism by which socially disadvantageous traits would be selected out was weakened. The upshot, Sterelny claims, is that humans retained their powers of instrumental reasoning but came to have preferences for things that did not enhance genetic fitness.

A different take comes from work by Samuelson and Swinkels (2006) and Rayo and Robson (forthcoming). They argue that the challenge is to explain why humans derive utility from *anything* other than biological reproduction itself. Food, sex and shelter, for example, obviously causally promote our fitness; however our desire for these goods is not purely instrumental. We desire tasty food as an end in itself, not simply because we know that consuming food will enhance our survival and hence our fitness. From an evolutionary viewpoint this seems odd. Given that biological fitness is what really matters, surely mother nature should have produced organisms who care non-instrumentally only about their fitness, and whose desires for "intermediate goods" like food and sex are purely instrumental? Yet modern humans are not like this. So according to this view, the challenge is not so much to explain why humans derive utility from things that are detrimental to fitness, but to explain why we derive utility from anything other than fitness itself.

The answer, according to the above authors, depends crucially on lack of information. Organisms are not born knowing the causal structure of the world, and can only learn some causal regularities by trial-and-error within their lifetime. Plausibly, the causal consequences for fitness of consuming different foodstuffs, having sex etc. are not something that our

ancestors could have learnt. If these causal consequences could be learnt, then mother nature could make each organism care only about fitness itself. After learning the relevant causal facts, organisms would then produce biologically optimal behaviour. But given that this is impossible, mother nature instead equips organisms with intrinsic (non-instrumental) desires for intermediate goods. Therefore humans have the utility functions they do precisely to compensate for their bounded rationality, i.e. the limitations on what can be learnt. This intriguing theory puts the connection between evolution, learning and rationality into a new perspective.

5. Conclusion

Traditionally the topic of rationality has been discussed without the benefit of a biological perspective, by philosophers, psychologists and economists. This traditional approach has undoubtedly yielded much interesting work. However as the brief survey above shows, a biological and in particular a Darwinian perspective offers the potential for new insights into the nature of rationality, both human and non-human, and suggests interesting new questions to ask. This is for three main reasons. Firstly, Darwinian fitness suggest a new normative yardstick – biological rationality – by which to evaluate beliefs and actions. Secondly, the cognitive capacities underlying rational thought and action are presumably evolved, raising the spectre of a Darwinian explanation of aspects of human rationality, and of our rational shortcomings. Thirdly, the science of evolutionary biology itself has drawn extensively on ideas from rational choice theory, suggesting a deep isomorphism between the fitness-maximizing paradigm of the former and the utility-maximizing paradigm of the latter. Each of these three topics is an ongoing field of enquiry.[6]

[6] This work was supported by the European Research Council Seventh Framework Program (FP7/2007–2013), ERC Grant agreement no. 295449.

References

Andrews, Kristin 2014. ‘‘Animal cognition’’, *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), E. N. Zalta (ed.) http://plato.stanford.edu/archives/fall2014/entries/cognition-animal/

Ariely, Dan 2008. *Predictably Irrational: the Hidden Forces that Shape our Decisions*, London: Harper Collins

Binmore, Ken 2005. *Natural Justice*. Oxford University Press

Charlseworth, Brian 1994. *Evolution in Age-Structured Populations*. Cambridge University Press

Chater, Nick 2012. ‘‘Building blocks of human decision making’’, in Hammerstein, Peter and Stevens, Jeffrey (eds.) *Evolution and the Mechanisms of Decision Making*, Cambridge MA: MIT Press, pp. 53–68

Clayton, Nicky, Emery, Nathan and Dickinson, Anthony 2006. ‘‘The rationality of animal memory’’, in Nudds and Hurley (eds.), pp. 197–216

Cosmides, Leda and Tooby, John 1994. ‘‘Better than rational: evolutionary psychology and the invisible hand,’’ *American Economic Review* 84: 327–332

Curry, Philip 2001. ‘‘Decision making under uncertainty and the evolution of interdependent preferences,’’ *Journal of Economic Theory* 98: 57–69

Danielson, Peter 2004. ‘‘Rationality and evolution’’, in Rawling, Piers and Mele, Alfred (eds.) *The Oxford Handbook of Rationality*. Oxford University Press, pp. 417–437

Davidson, Donald 1984. *Inquiries into Truth and Interpretation*. Oxford University Press.

Dennett, Daniel 1987. *The Intentional Stance*, Cambridge MA: MIT Press

Gardner, Andy and Grafen, Alan 2009. ‘‘Capturing the superorganism: a formal theory of group adaptation,’’ *Journal of Theoretical Biology* 22: 659–71

Gigerenzer, Gerd 2010. *Rationality for Mortals: How People Cope with Uncertainty*. Oxford University Press

Gigerenzer. Gerd and Selten, Reinhard 2001. *Bounded Rationality: the Adaptive Toolbox*. Cambridge MA: MIT Press

Gintis, Herbert 2009. *The Bounds of Reason*. Princeton University Press

Godfrey-Smith, Peter 1996. *Complexity and the Function of Mind in Nature*. Cambridge University Press

Grafen, Alan 1999. ‘‘Formal Darwinism, the individual-as-maximizing-agent analogy, and bet-hedging,’’ *Proceedings of the Royal Society B* 266: 799–803

Grafen, Alan 2006a. ‘‘Optimization of inclusive fitness,’’ *Journal of Theoretical Biology* 238: 541–63

Grafen, Alan 2006b. ‘‘A theory of Fisher’s reproductive value,’’ *Journal of Mathematical Biology* 53: 15–60

Grafen, Alan 2007. ‘‘The formal Darwinism project: a mid-term report,’’ *Journal of Evolutionary Biology* 20: 1243–54

Haig, David 2012. ‘‘The strategic gene,’’ *Biology and Philosophy* 27: 461–79

Hamilton, William 1964. ‘‘The genetical evolution of social behaviour I & II,’’ *Journal of Theoretical Biology* 7: 1–16, 17–52

Hammerstein, Peter and Stevens, Jeffrey 2014a. ‘‘Six reasons for invoking evolution in decision theory,’’ in Hammerstein and Stevens (eds.), pp. 1–20

Hammerstein, Peter and Stevens, Jeffrey (eds.) 2014b. *Evolution and the Mechanisms of Decision Making*. Cambridge MA: MIT Press

Houston, Alasdair and McNamara. John 1999. *Models of Adaptive Behaviour*. Cambridge University Press

Houston, Alasdair, McNamara, John and Steer, Mark 2007. ‘‘Do we expect natural selection to produce rational behaviour?’’ *Philosophical Transactions of the Royal Society B* 362: 1531–1543

Kacelnik, Alex 2006. ‘‘Meanings of rationality,’’ in Nudds and Hurley (eds), pp. 87–106

Kahneman, Daniel 2011. *Thinking Fast and Slow*. London: Penguin

Kahneman, Daniel and Tversky, Amos (eds.) 2000. *Choices, Values and Frames*. Cambridge University Press

Kahneman, Daniel, Slovic, Paul and Tversky, Amos 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press

Kennedy, John 1992. *The New Anthropomorphism*. Cambridge University Press.

Lewis, David 1981. ‘‘Causal decision theory,’’ *Australasian Journal of Philosophy* 59: 5–30

Martens, Johannes (forthcoming) ‘‘Hamilton meets causal decision theory,’’ *British Journal for the Philosophy of Science*

Maynard Smith, John 1974. ‘‘The theory of games and the evolution of animal conflicts,’’ *Journal of Theoretical Biology* 47: 209–221

Maynard Smith, John 1982. *Evolution and the Theory of Games*. Cambridge University Press.

McDowell, John. 1994. *Mind and World*. Harvard University Press.

Mylius, Sido and Diekmann, Odo 1995. ‘‘On evolutionarily stable life histories, optimization and the need to be specific about density dependence,’’ *Oikos* 74: 218–224

Nudds, Matthew and Hurley, Susan (eds.) 2006. *Animal Minds*. Oxford University Press

Okasha, Samir 2006. *Evolution and the Levels of Selection*. Oxford University Press

Okasha, Samir 2011. ‘‘Optimal choice in the face of risk: decision theory meets evolution,’’ *Philosophy of Science* 78: 83–104

Orr, Allen 2007. ‘‘Absolute fitness, relative fitness, and utility,’’ *Evolution* 61: 2997-3000

Quine, Willard 1969. ‘‘Epistemology naturalized,’’ in his *Ontological Relativity and Other Essays*. New York: Columbia University Press

Ramsey, Frank 1931. ‘‘Truth and probability,’’ in his *Foundations of Mathematics and Other Logical Essays*. New York: Harcourt

Rayo, Luis and Robson, Arthur (forthcoming) ‘‘Biology and the arguments of utility’’

Robson, Arthur 1996. ‘‘A biological basis for expected and non-expected utility,’’ *Journal of Economic Theory* 68: 397–424

Robson, Arthur and Samuelson, Larry 2011. ‘‘The evolutionary foundations of preferences,’’ in Bisin, Alberto and Jackson, Matthew (eds.) *Handbook of Social Economics*. Amsterdam: North-Holland, pp. 221–310

Samuelson, Larry and Swinkels, Jeroen 2006. ‘‘Information, evolution and utility,’’ *Theoretical Economics* 1: 119–42

Savage, Leonard 1954. *The Foundations of Statistics*. New York: Wiley

Seeley, Thomas 1996. *The Wisdom of the Hive*. Harvard University Press

Seeley, Thomas 2010. *Honey-Bee Democracy*. Princeton University Press

Skyrms, Brian 1996. *Evolution of the Social Contract*. Cambridge University Press

Sober, Elliott 1998. ‘‘Three differences between evolution and deliberation,’’ in Danielson, Peter (ed.) *Modelling Rationality, Morality and Evolution*. Oxford University Press, pp. 408–422

Stearns, Stephen 2000. ‘‘Daniel Bernoulli (1738): evolution and economics under risk,’’ *Journals of Biosciences* 25: 221–28

Stephens, Christopher 2001. ‘‘When is it selectively advantageous to have true beliefs?’’ *Philosophical Studies* 105: 161–189

Sterelny, Kim 2003. *Thought in a Hostile World*. Oxford: Blackwell

Sterelny, Kim 2012. ‘‘From fitness to utility,’’ in Okasha, Samir and Binmore, Ken (eds.) *Evolution and Rationality*. Cambridge University Press, pp. 246–73

Todd, Peter, Gigerenzer, Gerd and the ABC Research Group 2012. *Ecological Rationality: Intelligence in the World*. Oxford University Press

Weibull, Jorgen 1995. *Evolutionary Game Theory*. Cambridge MA: MIT Press

Wilson, David Sloan 2002. *Darwin’s Cathedral*. University of Chicago Press