

Niche construction, adaptive preferences, and the differences between fitness and utility

Armin W. Schulz

Received: 23 October 2013 / Accepted: 27 February 2014 / Published online: 6 March 2014
© Springer Science+Business Media Dordrecht 2014

Abstract A number of scholars have recently defended the claim that there is a close connection between the evolutionary biological notion of fitness and the economic notion of utility: both are said to refer to an organism's success in dealing with its environment, and both are said to play the same theoretical roles in their respective sciences. However, an analysis of two seemingly disparate but in fact structurally related phenomena—'niche construction' (the case where organisms change their environment to make it fit to their needs) and 'adaptive preferences' (the case where agents change their wants to make them fit to what the world has given them)—shows that one needs to be very careful about the postulation of this sort of fitness–utility connection. Specifically, I here use the analysis of these two phenomena to establish when connecting fitness and utility is and is not possible.

Keywords Evolutionary theory · Evolutionary economics · Natural selection · Rational choice

Introduction

A number of scholars have recently suggested that there is a close connection between the evolutionary biological notion of *fitness* and the economic notion of *utility* (Sterelny 2012a; Okasha 2007, 2011; Orr 2007; Rosenberg 2000, 134–135; Cooper 1987, 1989; Becker 1976): roughly, the thought here is that both evolutionary and economic success are premised on the maximization of what is fundamentally the same quantity—a quantity that happens to be labeled 'fitness' in evolutionary biology and 'utility' in economics. If correct, this supposition would be important, as it would suggest that insights from one domain can be carried over

A. W. Schulz (✉)

Department of Philosophy, University of Kansas, 3083 Wescoe Hall, Lawrence, KS 66045, USA
e-mail: awschulz@ku.edu

into the other, thus opening up new avenues for progress in the two subjects. However, as I try to make clear in this paper, the transition between utility and fitness is in fact not at all straightforward, and will only be compelling in a special set of cases.

To show this, I analyze the relationship between two phenomena: one evolutionary biological and one economic. On the one hand, there is the fact that organisms sometimes change their environments in order to make these environments better suited to what they happen to be, rather than change themselves to make themselves fit better to what their environments happen to be—a phenomenon that has become known as ‘niche construction’ (Lewontin 1982; Odling-Smee et al. 2003; Laland and Sterelny 2006). On the other hand, there is the fact that agents sometimes change their preferences about what they want to achieve so that they come to match what the world happens to be, rather than doing something to the world so that it comes to match what their preferences happen to be—a phenomenon that has become known as involving ‘adaptive preferences’ (Elster 1983; Bovens 1992; Sen 1995; Nussbaum 2000; Zimmerman 2003; Bruckner 2009; Hill 2009).¹ The reason for analyzing these two phenomena is that they turn out to be useful stepping stones in the formulation of a general criterion of when the transition between fitness and utility is defensible. (Moreover, in the course of this discussion, we can shed some new first-order light on the concepts of niche construction and adaptive preferences.)

The paper is structured as follows. In the first section, I discuss some of the linkages between fitness and utility that have recently been explored. In the following section, I analyze the logic of niche construction. In the section after that, I do the same for adaptive preferences. In the light of this discussion, I then return to the (supposed) parallels between fitness and utility. I conclude in the final section.

Fitness and utility

For some time now, there has been growing interest in building bridges between evolutionary biology and economics (Hodgson and Knudsen 2010; Vromen 2009; Hammerstein and Hagen 2005). For example, some authors have used models from gene-culture coevolutionary theory to analyze certain economic phenomena (Boyd and Richerson 1985, 2005; Nelson and Winter 1982; Hodgson and Knudsen 2010; Seabright 2010). Others have applied evolutionary game theory to human strategic decision making (Binmore 1998; Skyrms 1996; Samuelson and Binmore 2006; Alexander 2007). Yet others have applied economic theories of individual and social choice to non-human animals to predict and explain their decisions (Houston

¹ Note that the term ‘adaptive preferences’ is sometimes reserved for ‘bad’ (theoretically unacceptable) kinds of adaptive preference changes, and the term ‘character planning’ for ‘good’ (theoretically acceptable) ones (Elster 1983; Bovens 1992; Bruckner 2009). I return to this distinction below; for now it is just important to note that I shall be using the term ‘adaptive preferences’ in a wider sense that comprises both good and bad versions.

and McNamara 1999; List 2004; Conradt and List 2009; Okasha 2010; Noe and Voelkl 2013).

Importantly, moreover, a number of authors have also attempted to connect evolutionary biology and economics using a posited analogy between the concepts of *fitness* and *utility* (Cooper 1987; Stearns 2000; Gintis 2009; Alexander 2009; Okasha 2011; see also Hagen et al. 2012). On this approach, the idea is that these two core notions of evolutionary theory and economics match up so closely that insights in one of the two subjects can be transferred to the other—thereby opening up new avenues of progress there. It is this last way of attempting a synthesis between the two subjects that is at center stage here: while there is also much that can and should be said about other ways of connecting evolutionary biology and economics (Rosenberg 2000; Alexander 2007, 2009; Gruene-Yanoff 2011; Schulz 2013), the posited analogy between fitness and utility is sufficiently interesting—and sufficiently complex—to deserve a treatment of its own (see also Sterelny 2012a; Okasha 2011; Alexander 2007, 2009; Gruene-Yanoff 2011).²

To understand the fitness–utility connection better, it is best to start by dividing it into two claims:

- (a) In many cases of interest, fitness and utility are ordinally equivalent.
- (b) Fitness and utility function in the same way in evolutionary theory and economics.

Claim (a) says that, in many cases, assignments of fitness and utility values to a set of traits can reasonably be based on the same sorts of facts (e.g. the caloric values of different foodstuffs that can be obtained by different behavioral strategies), and will thus rank the traits in the same way. The general idea behind this claim is that, given the assumption that an organism’s preferences—represented by its utilities—play at least some role in determining what it does, then an organism that prefers what is biologically good for it is more likely to do those things that are fitness-enhancing. In turn, if what evolves is what is fitness enhancing and if these kinds of preferences are sufficiently heritable (both of which at least sometimes seems reasonable), they can be expected to sweep to fixation in the population. As Cooper (1987, 397) puts it: ‘Darwinian fitness and decision-theoretic utility are [here] seen as intimately related and almost the same thing. Specifically, expected utilities are conceived to play the role of hypothetical organismic fitness estimates, or internal constructs that tend to track fitness.’ For this reason, the same sort of facts can, at least sometimes, be used to ground ordinally equivalent fitness and utility assignments.

Claim (b) says that the use that evolutionary theory makes of assignments of fitness values is parallel to the use that economic theory makes of assignments of utility values. As Okasha (2011, 84) puts it: ‘My strategy will be to exploit an analogy between utility in rational choice and fitness in Darwinian evolution. (The former is the quantity that the rational agents try to maximize; the latter, that natural

² Note that the different ways of connecting evolutionary biology and economics are not completely separate, in that the same authors often pursue multiple such projects at the same time (see e.g. Trimmer et al. 2011; Alexander and Jason 2009). This, though, does not imply that there is no use in discussing the fitness–utility connection on its own—it is still based on a distinctive set of argumentative considerations.

selection tries to maximize.)' The idea behind this claim is that the same sort of constrained maximization analysis that is at the heart of much evolutionary biology also seems to be at heart of much economics (Grafen 1999; Hammerstein and Hagen 2005; Nelson and Winter 1982). So, on the one hand, at any given time, natural selection operates by selecting that trait F_i —out of the set of currently available traits F —that has the highest net fitness.³ In other words, it solves the equation

$$(NS) \text{Max}\{w(F_i)|F_i \in F\}$$

where $w(F_i)$ is the net fitness of having trait F_i .⁴ On the other hand, economic agency operates by choosing that action A_i —out of the set of currently available actions A —that has the highest net utility. In other words, it solves the equation

$$(EA) \text{Max}\{u(A_i)|A_i \in A\}$$

where $u(A_i)$ is the net utility of doing action A_i .⁵ In this way, the two sciences can seemingly be seen to deal with essentially identical processes.

To make clearer the kind of inferences allowed by the conjunction of claims (a) and (b), consider the following two examples.⁶ Firstly, note that there is much debate in economics over how to understand the attitudes towards risk displayed by people (see e.g. Gigerenzer and Selten 2001; Rabin and Thaler 2002; Orr 2007; see also Mithen 1990): should people be seen to be generally risk neutral, or should they be seen to be generally risk-averse (say)? Further, note that evolutionary biologists have found that, in general, the *variance* of a reproductive or survival strategy is a crucial factor determining its fitness (Gillespie 1977; Sober 2001). Given claims (a) and (b) above, this last point could then be taken to imply that economic agency—i.e. agency based on (EA)—should be seen to favor risk aversion (Orr 2007; Okasha 2007, 2011; Schulz 2008): given (a), low-variance (and therefore, low-risk) strategies should also be seen to have high utility (i.e. the utility of an expected value should be higher than the expected utility of the relevant outcome space), and given (b), what has high utility is all that matters for capturing people's decisions (at least as far as economics is concerned).⁷

³ In what follows, I shall switch between talking about the fitness of traits and that of organisms. While, in general, these two are not interchangeable (Sober 2000, chap. 3), for present purposes, conflating these is not problematic: I focus on only one trait per organism, so that the two come out to be the same here.

⁴ Note that this is different from the claim that natural selection maximizes the average fitness in the population. The latter claim is widely known to be true only in special cases (e.g. in the absence of frequency dependent selection or altruistic traits). What (NS) claims is merely that, *at any given time*, the fittest available trait is selected; which trait that is can change over time. Note also that (NS) should be seen to concern *net* fitness: there might be two (even opposing) effects on fitness—e.g. one via the individual level and one via the group level. Here, though, only the net effect matters (for more on this, see Okasha 2006).

⁵ Note that, as formulated above, (EA) fits most easily to a framework of certainty. However, the extension to the case of 'risk' is straightforward. See also below.

⁶ Note that these examples are just meant to provide a flavor of the issues; there is much more to be said about them.

⁷ Note that this leaves room for debate about exactly how the risk aversion is to be taken into account in economics—e.g. with non-linear utility functions or in some other way (see e.g. Kahneman and Tversky 1979). Note also that there is more to risk than variance—though variance is often taken to be a key feature and measure of risk (Mas-Colell et al. 1995). I thank Don Ross for useful discussion of this point.

Secondly, note that there is much debate in evolutionary biology over the extent to which natural selection is the most important evolutionary determinant—both in general and in particular cases (Orzack and Sober 1996; Godfrey-Smith 2001). For example, assume we are studying the evolutionary history of the food consumption decisions of red knots. Assume further it did turn out that these decisions now fit to the pattern that would be observed if these birds were to balance the utility of immediate food consumption with the utility of acquiring information about other available food sources in the vein of (EA) (van Gils et al. 2003; Olsson and Holmgren 1998; McNamara et al. 2012). Given (a) and (b) above, we could then conclude that, at least in this respect, the food consumption decisions of these birds are adaptations: given (a), these decisions have highest fitness, and given (b), this can be (best) explained by appealing to the workings of natural selection (Orzack and Sober 1996; Krebs and Davies 1997; Alcock 2009; Houston and McNamara 1999; Robson 2001; Trimmer et al. 2011).

How plausible, though, are claims (a) and (b)? Here, things are not straightforward. At first sight, it may appear that, of these two claims, it is claim (a) that requires the most defense: the reasoning behind this claim seems based on a model of cognitive evolution that leaves out much that might have an impact on which psychological dispositions evolve. In particular, at least for cultural species like our own, preferring what is popular or entertaining (though possibly maladaptive) can, at least at times, also be expected to spread in the population (Boyd and Richerson 2005; Sterelny 2012b). For this reason, the correlation between what an organism wants and what is biologically good for it should not always be thought to be high (see also Guth and Kliemt 1998; Chudek et al. 2013, pp. 436–437; Sterelny 2012a). By contrast, claim (b) might seem to be quite obviously true: as made clear earlier, apart from the change in the name of the variables, the two sciences seem to use fitness and utility functions in exactly the same way.

However, it turns out that things might actually be the reverse of this. While it is indeed true that claim (a) deserves a closer look, in a suitably restricted way, this claim is sufficiently plausible to help underwrite a fitness–utility connection in a number of cases of interest (see also Chudek et al. 2013, pp. 436–437; Sterelny 2012a; Boyd and Richerson 2005). In particular, the reasoning underlying this claim is sufficiently cogent to suggest that some correlation between fitness and utility rankings will often exist (though it need not be equally strong in all domains). By contrast, claim (b), despite its surface level plausibility, turns out to be highly problematic.

One reason for this is already quite widely known: while both (NS) and (EA) are optimizing processes, they are optimizing in a different way (Sober 1998).⁸ Specifically, evolution by natural selection is based on a ‘hill-climbing’ procedure—the optimal trait is found by successive, small improvements to an existing trait—whereas economic agency as laid out in (EA) is based on a globally

⁸ Another reason for doubting claim (b) has been suggested to lie in the fact that fitnesses are interpersonally comparable, while utilities are not (see e.g. Alexander 2007, 2009; Okasha 2011; Grueneyanoff 2011). As will be made clearer in what follows, I think this criticism indeed gets at important issues, but stops short of fully illuminating the differences between fitness and utility. Bringing this out is the aim of the rest of this paper.

optimizing procedure—the optimal action is found by determining *whichever* action has the highest (expected) utility. This matters, as it implies that evolution by natural selection need not reach the globally fittest trait—local fitness maximization is all it can directly accomplish.⁹

However, the worries with claim (b) in fact run deeper than this. For even if, in the case in question, local and global optimization lead to the same outcome (e.g. because the relevant functions are monotonically increasing—see also Sober (1998, 2008, chap. 3; Grafen 1999) understanding what maximizes fitness can be far from useful for understanding what maximizes utility (and vice versa).¹⁰ In fact, the reverse can be true: the quasi-identification of fitness and utility can obscure important issues in the two subjects, and thus make progress in evolutionary theory and economics harder, not easier. To bring this out, I want to suggest it is useful to consider the relationship among two interesting phenomena in evolutionary theory and economics: niche construction and adaptive preferences.

The reason why it is useful to compare these two phenomena is that, given (a) and (b) above, they seem to be structurally similar non-standard applications of (NS) and (EA) respectively.¹¹ In particular, both niche construction and adaptive preferences seem to be cases where the initially given fitness/utility values of the options (traits or actions) are being changed: instead of an organism maximizing its fitness/utility by adopting the option with the highest such value, the values of these options are changed such that the option the organism happens to have adopted ends up having the highest value. Importantly, though, while the two phenomena thus seem to be structurally related, their evaluation is quite different in the two sciences: in particular, most evolutionary biologists and philosophers see nothing in principle problematic about cases of niche construction, whereas most economists and philosophers find cases of adaptive preferences to be theoretically dubious. It is for this reason that comparing these two cases is so interesting: they appear to be structurally similar, but their evaluations are very different.

A final word concerning the argumentative strategy of this paper is in order. There can be no doubt that this strategy—comparing two theoretical notions using two non-standard phenomena across two different sciences—is somewhat circuitous. However, as I hope will become clearer at the end, in this case, the circuitous route in fact pays off.

⁹ It will reach the global optimum with probability approaching 1 only as the population size reaches infinity.

¹⁰ Note that, in what follows, I shall also not appeal to there being more to evolution than natural selection, or there being more to economic agency than utility maximization (see e.g. Gigerenzer and Selten 2001). The point here is that there are issues that create difficulties for attempts at connecting fitness and utility that are inherent to the structure of these notions—independently of their importance in evolutionary biology and economics.

¹¹ Interestingly, though, these two phenomena differ in that niche construction involves changes to the world (where the standard cases involve changes to the organism), whereas adaptive preferences involve changes to the agent itself (where the standard cases involve changes to the world). This, though, is not so relevant here.

Niche construction and fitness maximization

For a long time, the typical dynamics of evolution by natural selection have been conceived of as follows (Lewontin 1982, 1983; Brandon 1990, 6–9; Sterelny and Griffiths 1999, 268–292; Sober 2000; Godfrey-Smith 2009): the environment is structured in such a way as to pose some ‘adaptive problem’ to some population of organisms; there is variation in the population with regards to responses to this adaptive problem; finally, to the extent that these responses are heritable, the most adaptive among them get selected—i.e. spread in the population.

However, recently, a number of researchers have called for an emendation and revision of this picture: a greater emphasis should be placed on organisms changing their environments (Lewontin 1982; Odling-Smee 1988; Brandon 1990, 68–69; Godfrey-Smith 1996; Lewontin 2000; Sterelny 2003, 2012b; Odling-Smee et al. 2003; Laland and Sterelny 2006). In particular, instead of only seeing populations of organisms as changing themselves in the face of an adaptive problem, they should frequently also be seen as *dissolving* this problem by changing their environment in suitable ways. For example, a population of prairie dogs might be faced with frequent extreme fluctuations in temperature; to deal with this, instead of becoming more resistant to temperature changes, they could engineer their environment to be more stable in terms of temperature—for example, they could build and live in an expansive system of underground tunnels (Odling-Smee et al. 2003, 288–289). This idea has become known as ‘niche construction’: a population of organisms altering its own selective environment (Odling-Smee 1988; Brandon 1990; Odling-Smee et al. 2003; Laland and Sterelny 2006).¹²

In other words, defenders of the importance of niche construction suggest that, as far as evolutionary theory is concerned, there are two routes towards achieving high fitness (Odling-Smee et al. 2003, 375): either by making the organism’s needs match what the world is like or by making the world match what the organism needs. To make this clearer, reconsider the dynamics of evolution by natural selection in light of the idea of niche construction (Odling-Smee et al. 2003, 48–50, 239, 370–371, 2005; Okasha 2005; Sterelny 2005; Laland et al. 2005).¹³

Assume that there is variation in a population of organisms in relation to some set of traits F . Assume also that the organisms differ in terms of two *meta*-traits M_C and M_R : they either leave the fitness values of the elements of F as they are (the ‘conservative’ strategy M_C), or they engage in (possibly costly) activities that change these fitness values themselves—e.g. by altering the structure of the (physical or social) environment in certain ways (the ‘radical’ strategy M_R). Given this, when concentrating on natural selection, one should expect to evolve whatever

¹² Note that Odling-Smee et al. (2003) are using the term ‘niche construction’ widely to include all kinds of alterations of the environment—including migration into a new environment, metabolizing food, and photosynthesis. As will become clearer below, though, in this context, I will be using the notion in the narrower sense of changes to the environment *that are adaptive*. See also Okasha (2005), Sterelny (2005), Griffiths (2005), and Laland et al. (2005).

¹³ What follows is an extremely simplified and abstract reconstruction of these dynamics. For more detailed mathematical models of these sorts of cases, see Odling-Smee et al. (2003, chap. 3 and the relevant appendices).

feature is fittest—i.e. whichever feature brings organisms the greatest expected reproductive success (Sober 1984, 2008; Brandon 1990; Godfrey-Smith 2009). In particular, in line with (NS) laid out earlier, one should expect a ‘conservative’ solution to an adaptive problem (i.e. trait M_C) to evolve if and only if

$$w(F_i) > w'(F_j), \quad \text{for some } F_i \text{ and all } F_j \text{ in } F, \quad (1)$$

and a ‘radical’ solution (i.e. trait M_R) to evolve if and only if

$$w'(F_i) > w(F_j), \quad \text{for some } F_i \text{ and all } F_j \text{ in } F, \quad (2)$$

where $w(x)$ is the fitness of trait x in the unchanged environment, and $w'(x)$ is the fitness of trait x after it has been altered by M_R (Odling-Smee et al. 2003, 41–42, 236).¹⁴

In other words, on this reading, one should expect a population of organisms to pursue a ‘conservative’ solution (i.e. trait M_C to spread in the population) if it is better (in fitness terms) to leave the fitnesses of the F_k untouched and not engaging in—potentially costly—niche construction. Conversely, one would expect a population of organisms to pursue a ‘radical’ solution (i.e. trait M_R to spread in the population) if it is better (in fitness terms) to change the fitness values of the F_k and engaging in—potentially costly—niche construction (Odling-Smee et al. 2003, 301–302). For example, if a type of organism can achieve a higher fitness by remaining susceptible to environmental temperature fluctuations and taking certain steps to dampen these fluctuations (as compared both to remaining susceptible to environmental temperature fluctuations in an environment *with* temperature fluctuations and to changing internally so as to become more temperature fluctuation resistant), we would expect that the relevant activities of environmental temperature smoothing—such as tunnel building—will spread in the population (Odling-Smee et al. 2003, 41–42, 65, 116).¹⁵

¹⁴ Note that, technically, (1) and (2) would need to quantify over all the available ways of changing the environment [i.e. over all the different available ways of changing $w(x)$ into some different $w'(x)$]. For simplicity, I have left this out here. Note also that the situation here could also be captured by thinking of a matrix where the columns list all of the relevant traits, and the rows the different available fitness assignments to these traits, with the first row containing the current, unchanged fitness values. Then a conservative solution would consist of cases where the maximal value in this matrix is in the first row, and a radical solution in cases where the maximal value is in a row other than the first one. Finally, defining the fitnesses of M_R and M_C themselves is a little tricky, but a natural way of doing so would be by setting $w(M_C) = \max \{w(F_i)\}$ and $w(M_R) = \max \{w'(F_i)\}$.

¹⁵ There is an alternative way of conceptualizing niche construction (Dawkins 1982, chaps. 11–13; Odling-Smee et al. 2003, 30, 131–132, 191–192). On this alternative interpretation, one would make a distinction among the elements of F , with some being organism-focused (the F_k^O), and some environment-focused (the F_k^E). Given this, one should expect an organism-focused solution to an adaptive problem to evolve if and only if $w(F_i^O) > w(F_j^E)$, for some i and all j , and an environment-focused solution to evolve if and only if $w(F_i^E) > w(F_j^O)$, for some i and all j [where $w(x)$ is again the fitness of trait x]. However, for present purposes, this reading of niche construction is less useful than the one in the text; this is so for two reasons. Firstly, the interpretation in the text seems to be more in line with the main idea behind niche constructionism—namely, that the activities of organisms are both effects and causes of evolutionary processes (Lewontin 1982; Odling-Smee et al. 2003, 48, 112–113, 240; Sterelny 2005; Laland et al. 2005; Sterelny 2003, 148–149). Secondly, it is especially the interpretation in the text that is useful for clarifying the contrast to the phenomenon of adaptive preferences.

In short: one of the key elements of the niche constructionist approach is the claim that niche constructors can change the evolutionary process itself—they can affect the fitnesses of various traits (of their own or of other organisms) (Odling-Smee et al. 2003). Now, for present purposes, the key point to note concerning this claim is that, purely theoretically, it is entirely uncontroversial: given Eqs. (1) and (2), there is no reason intrinsic to evolutionary theory to think that niche construction—the evolution of ‘radical’ solutions to adaptive problems—will be any less important than the evolution of ‘conservative’ solutions (Odling-Smee 1988; Odling-Smee et al. 2003, chaps. 2 and 3; Sterelny 2003, 2005; Laland and Sterelny 2006). Of course, it might be that there are empirical reasons for why (1) is more often satisfied than (2); however, this will then be a contingent fact about the world, and not a principled matter of evolutionary theory. (Also, so far, no clear candidates for these empirical reasons have been found—in fact, the opposite is more likely to be true; see e.g. Odling-Smee et al. 2003, chap. 2). Hence, at least on the face of it, recent work in evolutionary ecology is right to stress the importance of ‘niche construction’: none of the ways of achieving a high organism/environment fit is suspect—as far as evolution by natural selection is concerned, the truth of Eq. (2) is just as unproblematic as that of (1). This is an important point to keep in mind for what follows below.

Adaptive preferences and utility maximization

From the point of view of standard economic theory, the broad outline of the way in which an agent solves a decision problem is as follows. The agent has certain views about what it would be *good* for the world to be like—i.e. she has certain desires—and she has certain views about what the world *is* like—i.e. she has certain beliefs. Her task then is to create as good a match as possible between what the world is like and what she wants it to be like, using the tools she believes are available to her (Jeffrey 1983; Nichols and Stich 2003; Hausman 2012).

In slightly more detail, according to the major theories of choice used in economics, it is assumed that (a) an agent has a set of preferences over what the world should be like (and which comprise both cognitive and conative components—see Hausman 2012); (b) these preferences satisfy a number of axioms (such as transitivity, completeness, etc.), (c) she acts on the basis of these preferences (Hausman 2012). According to the representation theorems of the theories of choice, it is then the case that the agent’s actions can be seen as those that maximize her (expected) utilities (Savage 1954; Luce and Raiffa 1957; Joyce 1999).¹⁶ In this way, standard economic theory sees agency as based on equation (EA) laid out

¹⁶ Note that there are three main ways of understanding the major theories of choice in economics (Hausman 2012; Schulz 2011; Ross 2005): firstly, we could see the theories as providing a (possibly idealized) description of the agent’s actual decision making processes; secondly, we could see them as predictive tools for an agent’s choices without descriptive intent as far as her actual psychology is concerned; finally, we could see them as providing a standard of rationality that choices or decisions have to satisfy. For present purposes, though, it is not necessary to single out a particular interpretation of these theories, so I shall formulate the argument of the paper in a way that remains neutral on this issue.

earlier: agents do what they most prefer, and this is equivalent to them maximizing their (expected) utilities.¹⁷

Importantly, from the point of view of standard economic theory, it is also envisioned that the agent solves a decision problem by picking *actions*: she chooses that action from the set of available ones that leads to outcomes most in line with her preferences. In other words, standard economic theory claims that the agent aims to make the world turn out most like she wants it to turn out. For example, if an agent contemplates buying a new car using her accumulated savings, standard economic theory says that she will choose that car and savings bundle that is highest in her preference ranking (Savage 1954; Hausman 2012).

In principle, however, this is not the only way for the agent to achieve a high degree of preference satisfaction: it would seem equally possible for her to simply change her preferences so that they match whatever the world happens to be like. For example, the agent could also make herself prefer keeping her current car and all of her savings to buying a new car and spending some of these savings—that way, she would also end up in a state where the world is the way she wants it to be.¹⁸ Scenarios of this type have become known as being based on ‘adaptive preferences’ (Elster 1983; Bovens 1992; Rickard 1995; Nussbaum 2000; Zimmerman 2003; Bruckner 2009; Hill 2009): they are situations in which agents change their wants to make them fit to whatever the state of the world happens to be, rather than change the state of the world to make it fit to their wants.¹⁹

Importantly, however, while both of these ways of achieving high desire satisfaction lead to an agent’s desires and the state of the world being in agreement, for many philosophers (and some economists), these two ways should not be seen to be theoretically on a par (Elster 1983; Bovens 1992; Sen 1995; Nussbaum 2000; Zimmerman 2003; Bruckner 2009). In fact, according to many scholars, only the first (world-based) way of achieving desire satisfaction is in line with economic theory; adaptive preferences are generally not even considered to be within the

¹⁷ This glosses over some contentious issues in economics—e.g. about the need for maximization (as opposed to merely satisficing); for more on this, see e.g. Gigerenzer and Selten (2001). For present purposes, though, I restrict myself to considering the standard framework. See also note 10 above.

¹⁸ More realistic instances of this sort of phenomenon include the fact that, in certain circumstances, oppressed people (e.g. women) seem to come to prefer a position of little political and social freedoms; equally, it appears that many consumers come to prefer whatever product they happen to own (Sen 1995; Nussbaum 2000; Kahneman et al. 1991). Apart from this, there is also much discussion in the medical literature about the fact that patients often seem to change the standards by which they assess their personal situation as that situation changes (a phenomenon known as ‘response shift’): see e.g. Sprangers and Schwartz (1999).

¹⁹ A related class of phenomena concerns cases of so-called ‘sour grapes’ (Elster 1983; Bovens 1992; Rickard 1995; Hill 2009): situations where agents change the way they think about the world to achieve a higher degree of desire satisfaction (the name refers to La Fontaine’s fable, in which a fox learns that some grapes he thinks look tasty are hanging too high for him to reach; he then changes his mind and claims that the grapes are—or at least look—sour). This case is different from that of adaptive preferences, in that ‘sour grapes’ involve changes in an agent’s *beliefs* (or *perceptions*) rather than changes in her *desires*. In general, cases of ‘sour grapes’ raise different issues from the ones that are being discussed here (Zimmerman 2003; Bruckner 2009; Hill 2009).

purview of the latter (Nussbaum 2000; Sen 1985, 1995).²⁰ What is less clear, though, is exactly why that is so—what, according to standard economic theory, makes accepting the state of the world and adjusting one's preferences to make them a better fit to that state so different from accepting one's preferences and making the state of the world a better fit to them?

To answer this question, it is useful to reconsider the case of adaptive preferences in the same framework that was employed earlier to describe cases of niche construction. Assume that an agent can do various actions A_k out of a set of actions A . Assume also that the agent has some preference ordering over these different A_k . Finally, assume that the agent can pursue two kinds of meta-strategies: she can leave her preference structure as it is (the 'conservative' strategy M_C), or she can change that preference structure in a given set of ways (the 'radical' strategy M_R).²¹ Given this, it seems that, according to (EA), it follows that we should expect M_C to be chosen if and only if

$$u(A_i) > u'(A_j), \quad \text{for some } A_i \text{ and all } A_j \text{ in } A, \quad (3)$$

and M_R to be chosen if and only if

$$u'(A_i) > u(A_j), \quad \text{for some } A_i \text{ and all } A_j \text{ in } A, \quad (4)$$

where $u(x)$ is the agent's utility assignment to action x before any possible changes in her preference structure, and $u'(x)$ is the utility assignment to action x after it has been changed by M_R .²² In other words, it seems that, in analogy with niche construction in evolutionary biology, it should be an upshot of the standard economic theories of choice that agents change their desires in certain ways (or at least act as if they did so) if that will lead to a higher degree of desire satisfaction, all things considered, than sticking to the desires they have and trying to find a way to make the world match them—and vice versa.

However, there is reason to think that this is in fact *false*: the standard economic theories of choice do *not* allow the derivation of (3) and (4) as suggested above. This is because, given how utilities enter into these theories, neither (3) nor (4) are meaningful within these theories at all. A fortiori, therefore, as they stand, these

²⁰ The existence of adaptive preferences is sometimes also seen as a *moral* issue—in particular, it might be claimed that the existence of adaptive preferences is one of the prime reasons why social policy should not (just) take an individual's preferences into account (Sen 1985, 1995; Nussbaum 2000, 2001; McKerlie 2007; Bykvist 2010; Hausman 2012). However, discussing this further is not necessary for present purposes.

²¹ Note that there is no clear analogue here to the organism-focused/environment-focused distinction mentioned in note 15: in economics, the actions an agent can do are not normally analyzed in a way that would make it possible to draw this distinction.

²² As before, Eqs. (3) and (4) should also quantify over the different feasible ways of changing an agent's preference structure; also as before, I shall leave this aside for simplicity. Note also that it need not be the case that all ways of changing the utility function are equally feasible: changing one's preferences may be difficult—e.g. in terms of time, concentration, and attention—so that only a few serious options actually exist at most times. Finally, as before, there is a matrix-based alternative way of conceiving this situation. See also note 14.

equations—and thus the adaptive preferences that they model—cannot be derived from these theories.²³

To see this, note that, within the standard economic theories of choice, utilities do not attach to actions outside of a specific evaluative structure: they are extremely weakly cardinal—or even merely ordinal—representations of a given preference profile, and there is no theory-inherent reason for using the same utility-scaling for different preference profiles. For this reason, utilities are not comparable across different preference structures: the different structures can be given very different kinds of utility representations. For example, in von Neumann–Morgenstern expected utility theory—probably the most common theory of choice used in economics (Mas-Colell et al. 1995)—²⁴ utility is only cardinal in first differences: if $u(x)$ is a function assigning utilities to the options in an agent's preference ordering, then all positive affine transformations of $u(x)$ are equally acceptable utility functions representing that same set of preferences (Luce and Raiffa 1957; Jeffrey 1983; Mas-Colell et al. 1995). Importantly, while this weak cardinality thus puts some slight limitations on the number of utility functions allowed to represent any given preference structure, it does not alter the fact that the number of possible utility functions is still so large as to make the comparison of utility values across different preference structures meaningless (the only thing that is comparable are ratios of utility differences). In particular, the utility functions used to represent different preference structures can still involve drastically different scalings and thus fail to provide any kind of basis on which to make comparison of utility values across preference structures.²⁵

²³ See Welsch (2005) and Hill (2009) for related accounts. Note, though, that Welsch's (2005) model concerns a case in which an agent changes the weights attached to various goals she pursues while keeping fixed the utility she achieves from these weight/goal combinations—which is slightly different from how the issues are set out here. Hill (2009) pursues a line more similar to the one suggested here, though he also does not draw out quite the same implications as is done here.

²⁴ Note also that, since circumstances of risk are ubiquitous, using a weaker theory than von Neumann–Morgenstern expected utility theory will not generally be possible either (and, at any rate, utility would then be even less narrowly circumscribed). I thank Don Ross for useful discussion of this point.

²⁵ So, for example, assuming that B is the gamble of getting A with probability 0.75 and D with probability 0.25, and C is the gamble of getting A with probability 0.25 and D with probability 0.75, then, according to von Neumann–Morgenstern expected utility theory, it must be that, for a utility function $u(x)$, $u(B) = 0.75u(A) + 0.25u(D)$ and $u(C) = 0.25u(A) + 0.75u(D)$ (this is due to the fact that, in von Neumann–Morgenstern expected utility theory, the utility of a gamble is assumed to be identical to the expectation of the utilities of the gamble's components). In turn, this implies that $u(B) - u(C) = 0.5[u(A) - u(D)]$; importantly, this will be so for all allowed rescalings of $u(x)$. However, we then still cannot say anything about the absolute values of A through D across the scalings—and hence, across preference structures. So, if an agent starts by preferring A to D [so has $u(A) > u(D)$] for some utility function $u(x)$, and then changes her mind and starts preferring D over A [so has $u'(D) > u'(A)$], for some utility function $u'(x)$, we know that her ranking of B and C must also flip [i.e. we know that $u(B) - u(C) > 0 > u'(B) - u'(C)$]. However, importantly, we cannot say that the utility she assigns to B before the change must be higher than the utility she assigns to C afterwards—after all, we could set $u(A) = 100$, $u(D) = 1$, giving $u(B) = 75.25$ and $u(C) = 25.75$, and $u'(D) = 1$, $u'(A) = 0$, giving $u'(B) = 0.25$ and $u'(C) = 0.75$. The only statements we can make here concern ratios of utility differences like this one: $[u(B) - u(C)]/[u(A) - u(D)] = [u'(B) - u'(C)]/[u'(A) - u'(D)]$. The latter, though, are too weak to capture cases of adaptive preferences, as they do not compare the utility values of the options themselves.

Importantly, this point about the non-comparability of utilities across preference structures holds whether these preference structures belong to different people or to different time slices of the same person: all that matters is that there are different preference profiles—i.e. different utility functions—²⁶ in play in the situation in question.²⁷ In turn, what this implies is that the ‘radical’ strategy above yields utility assignments that need not be comparable to their unchanged prior state: the utility values $u(x)$ and $u'(x)$ on either side of (3) and (4) do not need to have the same ‘scaling’. Hence, as such, these are not statements that can be made within the standard economic theories of choice—and they therefore cannot be seen as straightforward consequences of these theories.

Now, of course, it is possible to add something to the standard economic theories of choice to make the two utility functions comparable—there is nothing in these theories that precludes this (Harsanyi 1977; Bradley 2008; Binmore 2009).²⁸ One possibility here is to appeal to a ‘0 to 1’ rule (Hausman 1995): the least preferred option is given utility 0, the most preferred option utility 1, and all the rest are equally spaced out in between. Various other suchlike proposals exist: for example, to the extent that one has reason to think that the utility values in the particular case in question can be seen to represent some physiological quantity—such as strengths of pleasure or reward expectation (Schroeder 2004; Glimcher et al. 2005; Okasha 2011)—one might be able to restrict the permissible transformations of utility to a small set using this physiological quantity.

For present purposes, the key point to note concerning these additions to the standard economic theories of choice is that they are differentially plausible in different cases, and there will be some cases where no such additions are plausible (Hausman 1995). Since different utility functions can represent different things—moral values, physiological states, etc.—different rules for doing inter-/intrapersonal utility comparisons will be plausible in different circumstances.²⁹ For example, appealing to the ‘0 to 1’ rule might be quite plausible if the relevant preference profiles reflect an overall similarity in how many degrees of goodness

²⁶ In principle, differences in preference profiles could also stem from differences in beliefs; however, as made clear in note 19 above, this case raises different issues from the ones at stake here, and thus will not be further considered.

²⁷ The extent to which interpersonal comparisons of utility are similar to intrapersonal comparisons of utility is somewhat controversial; for discussion, see e.g. Hammond (1991), Broome (1991), Gibbard (1987), Griffin (1986, chap. 6); see also Binmore (2009). For present purposes, though, settling these issues is not necessary: no one in this literature thinks that utilities are *always* comparable in the intrapersonal case. This is all that matters here. For some useful recent discussions of the interpersonal comparability of utilities, see e.g. Sen (1979), Hausman (1995), Goldman (1995), Bradley (2008), and Binmore (2009).

²⁸ Note, though, that the common extensions of von Neumann–Morgenstern expected utility theory employed in practice—such as rank-dependent utility or the toleration of an agent choosing strictly dominated options—are *not* sufficient to achieve the needed kind of comparability (Diecidue and Wakker 2001). This is because these extensions only affect which kinds of transformations of a given utility function are equally acceptable as representations of a given preference ranking; however, they do not lead to the utility values themselves becoming comparable across preference structures. I thank Don Ross for useful discussion of this point.

²⁹ See Hausman (2012) for a defense of the view that preferences in economics can have many different sources. See also Ross (2005) for a related discussion.

there are seen to be, for this justifies giving the top and bottom option the same value in both cases. However, this assumption will not always be true: sometimes two different agents—or one agent before and after she has changed her preferences—can be assumed to evaluate the world so differently that what they prefer most or least differs in conative value—hence making the ‘0 to 1’ rule less compelling.

All of this matters for two reasons here. Firstly, it implies that economists or philosophers need not see adaptive preferences as being theoretically unproblematic. In particular, they are not forced to say that agents that change their preferences to make them match the world are maximizing their expected utilities as laid out in (EA)—for economic theory, as such, does not assess choices *across* different utility functions.

Secondly, though, economists and philosophers need not dismiss all cases of adaptive preferences as problematic either. In particular, they are now free to draw a distinction between ‘good’ and ‘bad’ kinds of adaptive preference changes, where the former are in line with economic theory—extended by principles of utility comparability—and the latter are not. Indeed, many of the classic discussions of adaptive preferences find a comfortable home in the present theoretical framework. For example, Elster’s (1983) claim that adaptive preferences are only defensible if they have been arrived at through conscious, intentional deliberation can be seen as an attempt to spell out when it is possible to make the relevant kinds of utility comparisons: when a preference is arrived at as the result of conscious, intentional deliberation, then that preference is comparable to its initial state (similar remarks can be made about other ways of justifying the defensibility of some adaptive preferences—e.g. those of Bovens 1992; Rickard 1995; Zimmerman 2003; Bruckner 2009).³⁰

Apart from their inherent interest, these two conclusions concerning the theoretical acceptability of adaptive preferences are important to note here, as they bring out some of the key differences between the notions of fitness and utility. In turn, these differences can be used to formulate a criterion of when fitness and utility can be connected to each other.

Fitness and utility revisited

In order to see why simple, straightforward quasi-identifications of fitness and utility will not work, note that there is a great difference between (1) & (2) and (3) & (4). While the former two can be derived perfectly validly from the theoretical framework of modern evolutionary biology, the same is not true for the latter two and the theoretical framework of modern economics.

Interestingly, it is also not too difficult to diagnose the *formal* reason for this asymmetry in what can be derived from the two theories. Fitness is normally seen as an inherently *probabilistic* notion—it concerns the expected number of offspring an

³⁰ Note the comparability of the relevant preferences does not mean that adaptive preferences *must* be in line with (EA): they would only be so if they did indeed maximize utilities in the case at hand.

organism has, maybe modulated by a function of the variance of that number (Gillespie 1977; Brandon 1990; Sober 2000, 2001). By contrast, utility, at least in its basic form, is not probabilistic in any way—it might combine with probabilities to represent an agent's preferences, but it itself is not probabilistic (Jeffrey 1983; Hausman 2012). This fact matters here, as it accounts for why fitness values typically can be changed and still remain comparable, while utility values typically cannot: from the get-go, there is more structure in the former—they are inherently less freely scalable than utility values—and hence they retain meaning even across different fitness assignments. In what follows, I will therefore say that fitnesses are *formally richer* than utilities.

Of course, this then raises the question of *why* utilities (at least inherently) are given a less formally rich representation than fitnesses. Why do economists accept, at least very often, a less strongly structured account of utility than evolutionary biologists do of fitness? In the main, there seem to be two reasons for this.³¹

Firstly, there is a difference in what economists want to do with the notion of utility as compared to what evolutionary biologists want to do with the notion of fitness. In particular, while it is one of the main goals of evolutionary biological analysis to determine which traits can be expected to evolve *in a given time period*, economists often just want to know which strategy has the highest utility.³² Importantly, in the latter case, we do *not* need to know how much 'better' an option is than some other option—something we definitely do need to know if it is our goal to calculate the speed of convergence towards the relevant stable state. For this reason, economists can be satisfied with a less constrained formal account of utility as compared to evolutionary biologists' account of fitness.

The second reason for why utilities have been given a less rich formal structure than fitnesses is that it is not so clear, for every case, what the appropriate richer structure would be for utilities (if any); equally, it is not clear that there is *one* relatively rich such structure that would be plausible for all cases. This is because—as noted earlier—utility differences can represent many different things, and it is not clear that one and the same formally rich structure would work for all of these different cases. So, for example, it may be that, when it is reasonable to see utilities as representing expectations of felt pleasures, they can be scaled using a measure with a fixed zero (e.g. the 'rest state' of the brain's pleasure centers) and a fixed unit (e.g. neurally least noticeable pleasure differences). By contrast, if utilities can be seen to represent (a subset of) the agent's largely stable moral concerns, a more appropriate measure might be the 0-to-1 rule mentioned earlier. In yet other cases, it might be completely unclear what the relevant measure is. Given this, it may well not be possible to opt for *one* richer account of utility.

Note that this is different in the case of fitness. The latter always captures the tendency or disposition of an organism to have offspring.³³ To be sure: there are

³¹ Note that I am here not trying to uncover the historical reasons for why a relatively 'unstructured' account of utility was adopted. I am merely trying to argue that there is something to be said in favor of this account (whatever historically led to its adoption).

³² Though not always: see e.g. Busemeyer and Townsend (1993).

³³ The nature of fitness is controversial, with some defending a subjectivist, reductive reading of this nature (Rosenberg 1994, 590–61). However, even these subjectivists about fitness would agree with what

many different ‘bases’ for these dispositions, but, given that they all concern reproductive success, it is not unreasonable to think that they can all be captured by the same sort of formal framework. Put differently: while there might be many different *reasons* for a given fitness difference in a number of organisms, what the fitness difference *represents* is always the same: a difference in the propensity (or some such) of the organism to have offspring. This, though, is not the case when it comes to utilities, where the same utility difference can *represent* very different things—e.g. differences in the expectation of felt pleasure or in the commitment to a non-personal cause (Hausman 2012). For this reason, it is much harder to come up with formally richer accounts of utility than it is for fitness.³⁴

Summarizing all of this, it thus becomes clear that fitness is a formally richer notion than utility because it is substantively narrower than the latter—it captures a smaller set of features of the relevant organism. This is important also because this formal richness is necessary for the work of evolutionary biologists. By contrast, utility is substantively wider than fitness, which makes it harder to give it a formally rich treatment. This, though, is not as problematic in economics, as, in many cases, a formally richer notion of utility is not needed.

Of course, this does not mean that this difference in the formal richness/substantive width of the two notions must always be relevant. It does not mean this, as some cases of evolutionary biological and economic analysis only involve one fitness or utility function and do not consider issues such as the speed of convergence to the relevant optimal value. The point to be noted is just that, if we want to appeal to a fitness–utility analogy in connecting work in evolutionary biology and economics, we must have reason to think that, in the case in question, comparisons across and within fitness/utility functions are either not necessary, or that all the utility functions in question rest on features of an agent’s moral outlook, physiology, etc. that permit the notion to have a formal richness that matches that of fitness functions.³⁵

In this way, we can generalize a point that has been made by others as well. In considering the potential uses of evolutionary game theory in economics, for example, it has sometimes been noted that this use is hampered by the fact that utilities, unlike fitnesses, are not, in general, interpersonally comparable (Alexander 2007, 2009; Okasha 2011; Gruene-Yanoff 2011). What the above discussion shows is that the problems for connecting evolutionary biology and economics using a fitness–utility analogy in fact run deeper than this. It is not just that fitness is

Footnote 33 continued

follows, but would merely add that these reproductive dispositions will, eventually, be replaced by precise statements of how many offspring each organism has. For more on fitness, see also Sober (2000, chap. 3, 2001).

³⁴ Note, though, that not all fitness comparisons need to be meaningful in biology either: in particular, those across species or taxa might not be. However, the latter failure of comparability would then not be due to the fact that there is no reason to see fitness as having the necessary probabilistic structure, but rather due to the fact that the relevant traits are typically not part of the same evolutionary population (Godfrey-Smith 2009; Sterelny and Griffiths 1999, chap. 12). I thank Samir Okasha and Elliott Sober for useful discussion of this issue.

³⁵ Another way of putting this point is that any ordinal equivalence between fitness and utility needs to hold not just synchronically and for the same agent, but also diachronically and across agents.

interpersonally comparable, and utility is not (in general); the real problem is that, in general, fitness is substantively narrower—and therefore formally less freely scalable—than utility. Because of this, connecting the two can fail even if issues of interpersonal comparability are excluded: even if we are only looking at one agent in isolation, we ought to be careful in applying a fitness–utility analogy.

An example based on human sexual strategies theory (Buss and Schmitt 1993; Gangestad and Simpson 2000) might make this point clearer.³⁶ Assume that, for one reason or another, it turned out that it is fittest for humans to be niche constructors with respect to their current sexual partners: in particular, assume that humans are most well off—in fitness terms—by changing the fitness landscape in such a way as to make staying with their current partner the most adaptive strategy (e.g. by getting married to them), rather than leaving this fitness landscape as it is, and searching for a higher quality partner elsewhere. Given this, what the present discussion shows is that we should not necessarily conclude that it is also utility maximizing for human agents to have adaptive preferences about their current sexual partner. Changing one's preferences so as to prefer being with one's current partner would only be utility maximizing in a meaningful sense if there were reason to think that the relevant utility assignments are formally rich enough to make this kind of comparison possible. Equally, just because it is not utility maximizing to make oneself prefer one's current mate (for one reason or another), it may still be *fitness maximizing*—in a niche constructionist way—to at least act as if one did so.

Putting all of this together, we can condense the preceding discussion into the following necessary and sufficient condition:

(Fitness-to-Utility) Natural selection (NS) and economic agency (EA) will lead to the same outcomes if and only if, in the case in question, (a) the given utility and fitness assignments are ordinally equivalent, (b) utility can be seen to have a formal structure that is rich enough to allow for comparisons across utility functions *or* only one utility/fitness function is appealed to,³⁷ and (c) there is no difference between global and local maximization.

Conclusion

I have tried to show that the notions of fitness and utility, despite some surface similarities, enter their respective theories in quite different ways. In particular, the former captures a more narrow set of features of the world, and is therefore formally richer than the latter. For this reason, some key phenomena—such as processes of niche construction and cases of adaptive preferences—can only be properly analyzed if these differences between the two notions are kept firmly in view: the formal richness (or lack thereof) of fitness and utility is an essential feature of some

³⁶ Note that what follows is again meant to be purely illustrative. For a critical examination of Sexual Strategies Theory, see e.g. Buller (2005) and Schulz (2010).

³⁷ The reason for the second disjunct in (b) is that some cases require merely ordinal comparisons across unchanged fitness or utility values. As noted earlier, these cases do not pose inherent difficulties for connecting fitnesses and utilities.

phenomena in evolutionary biology and economics and needs to be grasped as such for us to be able to fully comprehend these phenomena.

Importantly, establishing this is also interesting as it provides some further illumination of niche construction and adaptive preferences individually and in relation to each other. In particular, we come to see that the key issue that needs to be addressed when it comes to niche construction is *empirical* in nature (namely, the frequency and nature of activities that change the fitness values of an organism's traits), whereas the key issue that needs to be addressed when it comes to adaptive preferences is *theoretical* in nature (namely, the possibility of underwriting the formal richness of utility functions). Moreover, we learn that, if (a)–(c) in (Fitness to Utility) do hold, cases of niche construction and adaptive preferences can be seen to be theoretical mirror images of each other—a surprising result in and of itself.

Acknowledgments I would like to thank Samir Okasha, Luc Bovens, Elliott Sober, Don Ross, Mauro Rossi, an anonymous referee for this journal, and audiences at the universities of Edinburgh, Bern, and Bristol for useful comments on previous versions of this paper.

References

- Alcock J (2009) *Animal behavior: an evolutionary approach*, 9th edn. Sinauer Associates, Sunderland
- Alexander J (2007) *The structural evolution of morality*. Cambridge University Press, Cambridge
- Alexander J (2009) Evolutionary game theory. In: E. Zalta (ed) *The Stanford encyclopedia of philosophy* (Fall 2009 edition). <http://plato.stanford.edu/archives/fall2009/entries/game-evolutionary>
- Becker G (1976) Altruism, egoism, and genetic fitness: economics and sociobiology. *J Econ Lit* 14:817–826
- Binmore K (1998) *Game theory and the social contract*, vol II: Just playing. MIT Press, Cambridge
- Binmore K (2009) Interpersonal comparison of utility. In: Kincaid H (ed) *The Oxford handbook of philosophy of economics*. Oxford University Press, Oxford, pp 540–559
- Bovens L (1992) Sour grapes and character planning. *J Philos* 89:57–78
- Boyd R, Richerson P (1985) *Culture and the evolutionary process*. University of Chicago Press, Chicago
- Boyd R, Richerson P (2005) *The origin and evolution of cultures*. Oxford University Press, Oxford
- Bradley R (2008) Comparing evaluations. *Proc Aristot Soc* 108:85–100
- Brandon R (1990) *Adaptation and environment*. Princeton University Press, Princeton
- Broome J (1991) Utilitarian metaphysics? In: Elster J, Roemer J (eds) *Interpersonal comparisons of well-being*. Cambridge University Press, Cambridge, pp 70–98
- Bruckner D (2009) In defense of adaptive preferences. *Philos Stud* 142:307–324
- Buller D (2005) *Adapting minds*. MIT Press, Cambridge
- Busemeyer J, Townsend J (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol Rev* 100:432–459
- Buss D, Schmitt DP (1993) Sexual strategies theory: an evolutionary perspective on human mating. *Psychol Rev* 100:204–232
- Bykvist K (2010) Can unstable preferences provide a stable standard of well-being. *Econ Philos* 26:1–26
- Chudek M, Zhao W, Henrich J (2013) Culture-gene coevolution, large scale cooperation, and the shaping of human social psychology. In: Sterelny K, Joyce R, Calcott B, Fraser B (eds) *Cooperation and its evolution*. MIT Press, Cambridge, pp 425–457
- Conrath L, List C (2009) Group decisions in humans and animals: a survey. *Philos Trans R Soc B* 364:719–742
- Cooper W (1987) Decision theory as a branch of evolutionary theory: a biological derivation of the savage axioms. *Psychol Rev* 94:395–411
- Cooper W (1989) How evolutionary biology challenges the classical theory of rational choice. *Biol Philos* 4:457–481
- Dawkins R (1982) *The extended phenotype*. Oxford University Press, Oxford
- Diecidue E, Wakker P (2001) On the intuition of rank-dependent utility. *J Risk Uncertain* 23:281–298

- Elster J (1983) *Sour grapes*. Cambridge University Press, Cambridge
- Gangestad S, Simpson J (2000) The evolution of human mating: trade-offs and strategic pluralism. *Behav Brain Sci* 23:573–644
- Gibbard A (1987) Ordinal utilitarianism. In: Feiwel E (ed) *Arrow and the foundations of the theory of economic policy*. Macmillan, London, pp 135–153
- Gigerenzer G, Selten R (2001) *Bounded rationality: the adaptive toolbox*. The MIT Press, Cambridge
- Gillespie J (1977) Natural selection for variances in offspring numbers: a new evolutionary principle. *Am Nat* 111:1010–1014
- Gintis H (2009) *The bounds of reason*. Princeton University Press, Princeton
- Glimcher P, Dorris M, Bayer H (2005) Physiological utility theory and the neuroeconomics of choice. *Games Econ Behav* 52:213–256
- Godfrey-Smith P (1996) *Complexity and the function of mind in nature*. Cambridge University Press, Cambridge
- Godfrey-Smith P (2001) Three kinds of adaptationism. In: Orzack S, Sober E (eds) *Adaptationism and optimality*. Cambridge University Press, Cambridge, pp 335–357
- Godfrey-Smith P (2009) Darwinian populations and natural selection. Oxford University Press, Oxford
- Goldman A (1995) Simulation and interpersonal utility. *Ethics* 105:709–726
- Grafen A (1999) Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. *Proc R Soc B* 266:799–803
- Griffin J (1986) *Well being: its meaning, measurement, and moral importance*. Clarendon Press, Oxford
- Griffiths P (2005) Review of “niche construction”. *Biol Philos* 20:11–20
- Gruene-Yanoff T (2011) Evolutionary game theory, interpersonal comparisons and natural selection: a dilemma. *Philos Biol* 26:637–654
- Guth W, Kliemt H (1998) The indirect evolutionary approach: bridging the gap between rationality and adaptation. *Ration Soc* 10:377–399
- Hagen E, Chater N, Gallistel C, Houston A, Kacelnik A, Kalenscher T, Nettle D, Oppenheimer D, Stephens D (2012) Decision making: what can evolution do for us? In: Hammerstein P, Stevens J (eds) *Evolution and the mechanisms of decision making*. MIT Press, Cambridge, pp 97–126
- Hammerstein P, Hagen E (2005) The second wave of evolutionary economics in biology. *Trends Ecol Evol* 20:604–610
- Hammond P (1991) Interpersonal comparisons of utility: why and how they are and should be made. In: Elster J, Roemer J (eds) *Interpersonal comparisons of well-being*. Cambridge University Press, Cambridge, pp 200–255
- Harsanyi J (1977) *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge University Press, Cambridge
- Hausman D (1995) The impossibility of interpersonal utility comparisons. *Mind* 104:473–490
- Hausman D (2012) *Preference, value, choice, and welfare*. Cambridge University Press, Cambridge
- Hill B (2009) Three analyses of sour grapes. In: Gruene-Yanoff T, Hansson SO (eds) *Preference change*. Springer, Dordrecht, pp 27–56
- Hodgson G, Knudsen T (2010) *Darwin’s conjecture*. University of Chicago Press, Chicago
- Houston A, McNamara J (1999) *Models of adaptive behaviour*. Cambridge University Press, Cambridge
- Jeffrey R (1983) *The logic of decision*, 2nd edn. University of Chicago Press, Chicago
- Joyce J (1999) *The foundations of causal decision theory*. Cambridge University Press, Cambridge
- Kahneman D, Tversky A (1979) Prospect theory. *Econometrica* 47:313–327
- Kahneman D, Knetsch J, Thaler R (1991) Anomalies: the endowment effect, loss aversion, and status quo bias. *J Econ Perspect* 5:193–206
- Krebs J, Davies N (1997) *An introduction to behavioural ecology*, 4th edn. Wiley, New York
- Laland K, Sterelny K (2006) Seven reasons (not) to neglect niche construction. *Evolution* 60:1751–1762
- Laland K, Odling-Smee FJ, Feldman M (2005) On the breadth and significance of niche construction. *Biol Philos* 20:37–55
- Lewontin R (1982) *Organism and environment*. In: Plotkin HC (ed) *Learning, development and culture*. Wiley, New York
- Lewontin R (1983) *Gene, organism, and environment*. In: Bendall DS (ed) *Evolution from molecules to men*. Cambridge University Press, Cambridge
- Lewontin R (2000) *The triple helix*. Cambridge University Press, Cambridge
- List C (2004) Democracy in animal groups: a political science perspective. *Trends Ecol Evol* 19:1689–1699
- Luce D, Raiffa H (1957) *Games and decisions*. Wiley, New York

- Mas-Colell A, Whinston M, Green J (1995) *Microeconomic theory*. Oxford University Press, Oxford
- McKerlie D (2007) Rational choice, changes in values over time, and well-being. *Utilitas* 19:51–72
- McNamara J, Trimmer P, Houston A (2012) The ecological rationality of state-dependent valuation. *Psychol Rev* 119:114–119
- Mithen S (1990) *Thoughtful foragers*. Cambridge University Press, Cambridge
- Nelson R, Winter S (1982) *An evolutionary theory of economic change*. Belknap Press, Cambridge
- Nichols S, Stich S (2003) *Mindreading*. Oxford University Press, Oxford
- Noe R, Voelkl B (2013) Cooperation and biological markets: the power of partner choice. In: Sterelny K, Joyce R, Calcott B, Fraser B (eds) *Cooperation and its evolution*. MIT Press, Cambridge, pp 131–151
- Nussbaum M (2000) *Women and human development*. Cambridge University Press, Cambridge
- Nussbaum M (2001) Symposium on Amartya Sen's philosophy: 5 adaptive preferences and women's options. *Econ Philos* 17:67–88
- Odling-Smee K (1988) Niche constructing phenotypes. In: Plotkin HC (ed) *The role of behavior in evolution*. MIT Press, Cambridge, pp 73–132
- Odling-Smee FJ, Laland K, Feldman M (2003) *Niche construction*. Princeton University Press, Princeton
- Okasha S (2005) On niche construction and extended evolutionary theory. *Biol Philos* 20:1–10
- Okasha S (2006) *Evolution and the levels of selection*. Oxford University Press, Oxford
- Okasha S (2007) Rational choice, risk aversion, and evolution. *J Philo* 104:217–235
- Okasha S (2010) Individuals, groups, fitness and utility: levels of selection meets social choice theory. *Biol Philos* 24:561–584
- Okasha S (2011) Optimal choice in the face of risk: decision theory meets evolution. *Philos Sci* 78:83–104
- Olsson O, Holmgren N (1998) The survival-rate-maximizing policy for Bayesian foragers: wait for good news. *Behav Ecol* 9:345–353
- Orr HA (2007) Absolute fitness, relative fitness, and utility. *Evolution* 61:2997–3000
- Orzack SH, Sober E (1996) How to formulate and test adaptationism. *Am Nat* 148:202–210
- Rabin M, Thaler R (2002) Risk Aversion. *J Econ Perspect* 15:219–232
- Rickard M (1995) Sour grapes, rational desires and objective consequentialism. *Philos Stud* 80:279–303
- Robson A (2001) Why would nature give individuals utility functions? *J Polit Econ* 109:900–914
- Rosenberg A (1994) *Instrumental biology*. University of Chicago Press, Chicago
- Rosenberg A (2000) The biological justification of ethics: a best-case scenario. In: *Darwinism in philosophy, social science, and policy*. Cambridge University Press, Cambridge, pp 118–136
- Ross D (2005) *Economic theory and cognitive science: microexplanation*. MIT Press, Cambridge
- Samuelson L, Binmore K (2006) The evolution of focal points. *Games Econ Behav* 55:21–42
- Savage L (1954) *The foundations of statistics*. Wiley, New York
- Schroeder T (2004) *Three faces of desire*. Oxford University Press, Oxford
- Schulz A (2008) Risky business: evolutionary theory and human attitudes towards risk—a reply to Okasha. *J Philos* 105:156–165
- Schulz A (2010) It takes two: sexual strategies and game theory. *Stud Hist Philos Biol Biomed Sci* 41:41–49
- Schulz A (2011) Gigerenzer's evolutionary arguments against rational choice theory: an assessment. *Philos Sci* 78:1272–1282
- Schulz A (2013) Beyond the hype. *Philos Soc Sci* 43:46–72
- Seabright P (2010) *The company of strangers*. Princeton University Press, Princeton
- Sen A (1979) Interpersonal comparisons of welfare. In: Boskin M (ed) *Economics and human welfare*. Academic Press, New York
- Sen A (1985) *Commodities and capabilities*. North Holland Publishing, Amsterdam
- Sen A (1995) Gender inequality and theories of justice. In: Nussbaum M, Glover J (eds) *Women, culture, and development*. Clarendon Press, Oxford, pp 259–273
- Skyrms B (1996) *Evolution and the social contract*. Cambridge University Press, Cambridge
- Sober E (1984) *The nature of selection*. University of Chicago Press, Chicago
- Sober E (1998) Three differences between deliberation and evolution. In: Danielson P (ed) *Modeling rationality, morality, and evolution*. Oxford University Press, Oxford, pp 408–422
- Sober E (2000) *Philosophy of biology*, 2nd edn. Westview Press, Boulder
- Sober E (2001) The two faces of fitness. In: Singh R, Paul D, Krimbas C, Beatty J (eds) *Thinking about evolution: historical, philosophical, and political perspectives*. Cambridge University Press, Cambridge, pp 309–321

- Sober E (2008) *Evidence and evolution*. Cambridge University Press, Cambridge
- Sprangers M, Schwartz C (1999) Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med* 48:1507–1515
- Stearns S (2000) Daniel Bernoulli (1738): Evolution and economics under risk. *J Biosci* 25:221–228
- Sterelny K (2003) *Thought in a hostile world*. Blackwell, Oxford
- Sterelny K (2005) Made by each other: organisms and their environment. *Biol Philos* 20:21–36
- Sterelny K (2012a) From fitness to utility. In: Binmore K, Okasha S (eds) *Evolution and rationality: decision, cooperation, and strategic behaviour*. Cambridge University Press, Cambridge, pp 246–273
- Sterelny K (2012b) *The evolved apprentice*. MIT Press, Cambridge
- Sterelny K, Griffiths P (1999) *Sex and death*. University of Chicago Press, Chicago
- Trimmer P, Houston A, Marshall J, Mendl M, Paul E, McNamara J (2011) Decision-making under uncertainty: biases and Bayesians. *Anim Cogn* 14:465–476
- van Gils J, Schenk I, Bos O, Piersma T (2003) Incompletely informed shorebirds that face a digestive constraint maximize net energy gain when exploiting patches. *Am Nat* 161:777–793
- Vromen J (2009) Advancing evolutionary explanations in economics: the limited usefulness of tinbergen's four-question classification. In: Kincaid H (ed) *The Oxford handbook of philosophy of economics*. Oxford University Press, Oxford, pp 337–368
- Welsch H (2005) Adaptation of tastes to constraints. *Theor Decis* 57:379–395
- Zimmerman D (2003) Sour grapes, self-abnegation, and character building. *Monist* 86:220–241