The Benefits of Rule Following:

A New Account of the Evolution of Desires

Armin Schulz

Department of Philosophy, Logic, and Scientific Method

London School of Economics and Political Science

Houghton St.

London WC2A 2AE

UK

a.w.schulz@lse.ac.uk

(0044) 753-105-3158

Abstract

A key component of much current research in behavioral ecology, cognitive science, and economics is a model of the mind at least partly based on beliefs and desires. However, despite this prevalence, there are still many open questions concerning both the structure and the applicability of this model. This is especially so when it comes to its 'desire' part: in particular, it is not yet entirely clear when and why we should expect organisms to be desire-based – understood so as to imply that they consult explicit tokenings of what they ought to do – as opposed to be drive-based – understood so as to imply that they react to the world using behavioral reflexes. In this paper, I present the beginnings of an answer to this question. To do this, I start by showing that an influential recent attempt to address these issues – due to Kim Sterelny – fails to be fully successful, as it does not make sufficiently clear what the relative benefits and disadvantages of drive-based and desire-based cognitive architectures are. I then present an alternative account of this matter based on the idea that organisms that can follow explicit behavioral rules (i.e. which have desires) avoid having to memorize a large set of state of the world-action connections – which can (though need not) be adaptive. Finally, I apply this account to the question of what the cognitive value of mental representations should be seen to be; here, I conclude that – contrary to some recent claims – relying on mental representations can make decision making easier, not harder, but also that – in line with these recent claims – whether it does so depends on the details of the case.

The Benefits of Rule Following:

A New Account of the Evolution of Desires

I. Introduction

A model of the mind underlying much work in (cognitive and evolutionary) ethology, psychology, economics, and philosophy is that of a belief / desire psychology (see e.g. Carruthers, 2006; Hausman, 2012; Nichols & Stich, 2003; Bekoff et al., 2002; Heyes & Huber, 2000; Allen & Bekoff, 1997; see also Allen, 2004): at least some of an organism's actions are assumed to be caused by what it thinks the world is like (often referred to as its 'beliefs'), and by what it thinks it ought to do in the situation it is in (often referred to as its 'desires') – i.e. by consulting the explicit content of certain types of mental states.[1] Given this, the study of this belief / desire model of the mind obviously has great theoretical and practical importance. Now, as it happens, one of the most fruitful recent approaches to this study has been an evolutionary one: various authors have investigated the reasons for why something like beliefs and desires might have evolved, and what this might imply about how these mental states work in determining behavior (see e.g. Sober, 1994; Godfrey-Smith, 1996; Sterelny, 2003).[2]

[1] Note that the claim here is not that all of an organism's actions are determined by content-bearing mental states like beliefs and desires – even if that organism is a human being (see also note 34 below). The claim is just that, at least for some organisms, some of their actions are.

[2] Note that the concepts of 'belief' and 'desire' involved in this context are technical notions whose relationship to the ordinary, folk psychological concepts with the same name is an open question (see e.g. Sterelny, 2003, Papineau, 2004, and Stich, 2004). However, note also that I do assume – in line with most of the discussions of this topic and possibly also folk psychology – that desires are content-bearing states (i.e. that they are representational). This assumption is not greatly restrictive, however, since even accounts that deny it (such as those of Anscombe, 1957, and Smith, 1987) may well have room for states that represent what the organism is to do, even though they might not call them desires (see Schroeder, 2009, and Railton, 2012 for more on different views of what desires express). Relatedly, I here do not engage in the dispute over whether there are any structural differences between beliefs and desires (see e.g. Lewis, 1988, 1996.). My question is: why should an organism use representations of what it is to do to determine its actions; whether this is to be cashed out as involving normative beliefs (or the like) or desires is something that I can leave open here. However, to make the exposition clearer – and because of existing terminological precedent (see below in note 10) – I continue to refer to these representations as 'desires'; readers with different views can feel free to replace this with a different term, though (e.g. 'normative beliefs').

However, there is one highly peculiar feature about this evolutionary approach that needs to be noted: it is largely focused on *beliefs* only – *desires* are (with a few exceptions) hardly ever explicitly considered.[3] Given the obvious importance of *both* beliefs and desires for the belief / desire model of the mind, this therefore raises two questions: (a) why did desires evolve? (b) what implications does an answer to (a) have for our understanding of the belief / desire model of the mind? In this paper, I present an account that is meant to go some ways towards answering these questions.

To do this, I begin by critically discussing Sterelny's (2003) arguments for why something like desires might have evolved in section II. I then present my own account of the evolution of desires in section III. I consider the implications of this account in section IV. I conclude in section V.

II. Sterelny's Account of the Evolution of Desires

In order to present my account of the evolution of desires most clearly, it is best to begin by assessing an influential recent account in the literature: that of Sterelny (2003, chap. 5).[4]

[3] Note, though, that there is some work in the economic literature on the evolution of 'preferences': see e.g. Robson (2001), Samuelson (2001), Guth (1995); Robson & Samuelson (2008). However, the aim of this work is quite different from the approach taken here. For example, Robson (2001) tries to determine when motivational structures that allow for learning are more adaptive than ones that are 'hardwired' – which, though, cross-cuts the issues of importance here. Similarly, Samuelson (2001) tries to determine when complexity constraints lead organisms to move away from a purely associationist motivational architecture; his argument too, though, is based on the idea that associationist architectures must be innate and unchangeable. I return to some of these issues below.

[4] Predecessors of this account can be found in Sterelny (1999, 2001). For somewhat related accounts, see also Kirsh (1996), McFarland (1996), and Dickinson & Balleine (2000). I focus on that of Sterelny (2003), as it is more conducive to discussing the issues of interest here: the evolution of desires by themselves, and the implications of the latter for the importance of mental representations in cognition. Another account of the evolution of desires that ought to be mentioned here is in Millikan (2002). Millikan suggests that organisms evolutionarily start out by relying on representations that are both descriptive and directive – 'pushmi-pullyu' representations – and that these two representational functions then, over time, get divided into dedicated mental states (i.e. beliefs and desires). In particular, she suggests that organisms that can represent their goal states and the state of the world separately from each other can (a) monitor when they have achieved what they set out to achieve, and (b) find new ways of achieving their goals – both of which may be adaptive. Now, it may appear that this account is drastically different from the ones defended by Sterelny and me. However, for two reasons, I think it is in fact more congenial to the latter than it might at first seem. Firstly, the fact that Millikan is concerned with the evolution of organisms that rely

Sterelny's account of the evolution of desires has two elements, one more negative in outlook and one more positive. Consider these two elements in turn.

The more negative part of Sterelny's argument centers on the contention that, since beliefs and desires need to be seen as selective responses to very different environments, we cannot look to the reasons for why beliefs have evolved to inform us about the reasons for why desires have evolved (see e.g. Sterelny, 2003, pp. 78-81). In particular, Sterelny claims that, whereas beliefs have the (evolutionary) function to carry information about the *external* environment of the organism – i.e. the state of the world – desires have the (evolutionary) function to carry information about the *internal* environment of the organism – i.e. the state of its needs.[5] However, he also thinks that the internal environment of the organism, unlike its external environment, is not epistemically complex: the organism's needs do not disguise or hide themselves, but signal their presence truthfully. This matters, as Sterelny thinks that environmental epistemic complexity is the prime driver of the evolution of beliefs: according to Sterelny (2003, chaps. 3-4), the ability to form beliefs is a selective response to environmental

---

on inner representations of goal states, whereas I am concerned with the evolution of organisms that rely on inner representations of what they are to do (it is not entirely clear which of these Sterelny is concerned with; see also note 5) does not seem greatly relevant, as these two representations are closely related. In particular, in representing what it is to do, an organism can also be seen to be (implicitly) representing the goal of its behavior – i.e. to bring about states of the world that constitute following the relevant rule. (The converse is also true, in that an organism representing a goal state can at least implicitly be seen as also representing the rule 'act so as to bring about this state' – though it does not necessarily specify a particular way in which this state is to be brought about.). In turn, this makes Millikan's account and mine quite easily inter-translatable in this regard. Secondly, the difference in the starting points (pushmi-pullyu representations versus non-representational reflexes) does not seem to be greatly problematic either, since Sterelny's account and mine should apply to either of these starting points. Millikan argues that pushmi-pullyu representations split into beliefs and desires since the former cannot be used in quite the same way that beliefs and desires can (they are in a different 'representational format'). This implies, though, that Millikan's account also depends on identifying the benefits that beliefs and desires bring on their own – just like Sterelny's account and mine do. That said, it is true that the reasons she puts forwards for why beliefs and desires should (sometimes) be expected to evolve are different from the reasons put forward here. In turn, this seems mostly due to the fact that Millikan sees beliefs and desires as necessarily coevolving, whereas (as made clearer below), I think it is useful to follow Sterelny in seeing the evolution of these two as, at least in principle, separate. Still, even here, there may be more common ground between Millikan's treatment and the ones laid out by Sterelny and me (see e.g. note 7); however, discussing this in detail has to be postponed to a different occasion.

[5] It thus seems that Sterelny (2003) does not distinguish between the belief that one is dehydrated and the desire to find water. As noted earlier (see note 2), while there is room to question this conflation, I will not do so here. I thank an anonymous referee for useful discussion of this point.

epistemic complexity. Hence, he concludes, desires cannot have evolved for the reasons that beliefs have evolved: they cannot be devices to deal with epistemic complexity, as this kind of complexity was not present in the selective environments relevant to them.

Now, while there are several places where this negative argument can be questioned, for present purposes, it useful to at least accept its conclusion. This is so for two reasons. Firstly, I think Sterelny is right to see beliefs and desires as traits that are, at least in principle, dissociable. This is because I think it should be left open as a possibility that an organism has beliefs, but that it uses these beliefs only to trigger a behavioral response in a way that is not mediated by the content of some other kind of mental state – i.e. that it lacks desires.[6] Equally, I think it should be left open as a possibility that an organism reacts to the world just through the states of its basic input systems (e.g. its retinal impressions) – i.e. that it lacks beliefs – but that the way in which it uses these inputs states to decide on what to do involves desires (e.g. the organism might rely on representations like 'the thing to do is to make *these kinds* of retinal impressions larger'). Secondly and relatedly, I think Sterelny is right to see beliefs and desires as playing sufficiently different causal roles in an organism's cognitive architecture to make it plausible that their evolution follows, at least in principle, separate paths.[7] In particular, as I also try to make clearer in what follows, there may be cases where it is adaptive for an organism to form beliefs, but where it is not adaptive for it to rely on desires. In this way, the rest of this paper can also be seen

[6] Millikan (1995, 2002), might be tempted to refer to representations that directly trigger a behavioral response as 'pushmi-pullyu' representations. However, the plausibility of this would seem to depend on the selective history of the mechanisms for generating these representations: if the organism relied on these representations not for bringing about a particular form of behavior, but for other reasons – such as saving energy or computational power – they might be better referred to as beliefs; see also Millikan (2002, chap. 14). However, discussing this in detail has to be postponed to another occasion.

[7] Of course, all of this is consistent with it being the case that organisms that have desires typically also have beliefs. However, this would then be so for contingent reasons: for example, it may be that if an organism already has evolved the machinery to rely on belief-like content when deciding what to do, it will be easier for it to also do so when it comes to desire-like content. In principle, though, the two can come apart.

as an attempt to show the usefulness of asking for separate accounts of the evolution of beliefs and desires.

Given this, consider the more positive element of Sterelny's account of the evolution of desires. Here, Sterelny begins by claiming that the major alternative to making decisions using desires is to do so using drives. Since this distinction is underlying the rest of the discussion here, it is very important to get clearer on it.[8]

Drives – as they are understood here – are states of an organism that dispose it to act in certain ways, but which do not rely on any kind of explicit content in doing so. Paradigmatically, they are based on look-up tables: they consist in mappings between states of the world and actions. Note that they still might *have* content (e.g. they might have been selected to reliably indicate the state of the organism's needs: see Millikan, 1984, 1990, 1995; Papineau, 1987) – the point is just that this content is not (directly) *used* when they determine the organism's behaviors; the content is (at most) an implicit by-product of the way they function.[9] By contrast, *desires* – as they are understood here – are states of an organism that dispose it to act in certain ways, and which do so precisely through this kind of explicit content: they express what the organism ought to do, and guide the organism's behavior by means of this expression. In other words, the difference between drive-based and desire-based organisms is that, in the latter case, mental content – and specifically, the consultation of a behavioral rule – mediates between the organism's assessment of what the world is like and its actions.[10]

[8] For more on the contrast between drives and desires, see also Kirsh (1996), McFarland (1996), and Spier & McFarland (1998).

[9] For an analogy, consider a person who is driving towards downtown, and who gets confused by a flashing sign saying 'For downtown, use right lane'; assume further that her temporary state of confusion causes her to veer into the right lane. In this case, the sign clearly has content, and the person is reacting to the sign, but she is not reacting to the sign by using its content (it can be assumed) – she might not have actually read what it said.

[10] Note that this is in line with many common views of what desires are meant to do: see e.g. Millikan (1995, 2002), Schroeder (2004, 2009), Railton (2012). Note also that this distinction between relying on mental states with explicit and (at most) implicit content is different from that between behavior based on stimulus-response (SR) and behavior

Note that what it means for a mental state to have content – i.e. to be representational – is tricky and has provoked considerable discussion: it might involve the deployment of a mental state that has a specific selective, ontogenetic, or causal history, or of one that resembles a state of the world in a particular way, or something else altogether (see e.g. Dretske, 1981, 1988; Millikan, 1984, 1990, 1996; Stampe, 1986; Papineau, 1987; Davidson, 1987; Fodor, 1990; Sterelny, 1991; Prinz, 2002, Neander, 2012). Fortunately, for present purposes, a detailed treatment of this issue is not necessary: the question asked here is '*why* would an organism rely on behavioral rules – i.e. mental content – to make decisions?'. This question is different from – and can be answered independently from – the question of what, exactly, it *means* (philosophically) to rely on behavioral rules / mental content: the former question presupposes that, somehow, it is possible to answer the latter question, but it does not depend on any particular such answer.[11]

The one thing that does matter here is that relying on a desire that X requires more than just acting *as if* X were thing to do: an organism that relies on a desire that X, in some sense and in some way, must mentally token that *X is the thing to do* and use this token to determine how to behave. A classic way of marking this distinction is by distinguishing between 'following a rule' and 'acting in accordance with a rule' (for more on this, see e.g. Kripke, 1982; Davidson, 1982;

based on operant reward (OR) learning (see e.g. Mackintosh, 1983): either of the latter two could be subserved, in the organism, by either drives or desires. Equally, the distinction between drives and desires is not a matter of whether the decision making system in the organism can be modeled as a signaling system in the sense of Skyrms (2010): if this sort of model is found appropriate, the issues at stake here would remain, as they concern what the receiver systems in the organism *do* with the signal about the state of the world they receive from the sender systems. Do they simply map it to a given action, or do they employ it as an input to a calculation about what the appropriate action is? Finally, note that it is not presumed that the explicit tokenings underlying desires have to be conscious; see also Clark (1991, 1992) for more on the distinction between explicit and implicit content.

[11] In other words, I am here not providing a philosophical account of content or meaning. In particular, I am not trying to solve the 'rule following problem': how *could* an organism rely on content-bearing mental states (rules) to determine its behavior? Does it need to be able to follow rules in order to interpret the mental injunction to follow some particular behavioral rule – thus leading to an infinite regress? Is there an alternative way of understanding and following rules (for discussion of this problem, see e.g. Kripke, 1982; Boghossian, 1989; Millikan, 1990; Schulz, 2009; Cheng, 2011)? Instead, I here simply assume that a solution to this problem can, somehow, be formulated (for an influential attempt that is congenial to the treatment laid out here, see Millikan, 1990).

Mele, 1987; Harre, 2002). The latter is something that drive-based organisms can do, too: by following a mapping of states of the world to action, they *implicitly* act in accordance with a rule – in fact, they might act in accordance with the same rule that their desire-based competitors do. However, they do not act the way they do because they have specifically *consulted* this rule – it is just that their actions implicitly *fit to* this rule. This is comparable to the different ways of determining the value (to a given degree of precision) of a function at a certain input: to do this, one can either consult an extensive mapping of inputs to outputs (as in S = {<1, 1>, <2, 1.4>, <3, 1.7>, <4, 2>, ...}), or one can consider the function itself (as in f(x) = $\sqrt{x}$, rounded to one decimal place). Of course, in using the mapping, one *is* calculating the relevant function; however, the point to note is that, when doing so, one does not *consult* this function, but one merely goes through an *enumeration* of all the solutions of that function. The same is going on here: drive-based organisms might well act in accordance with the same rule that desire-based organisms do – however, they do not *follow* this rule, but merely act in *accordance* with it.[12]

Note that this distinction between *following a rule* and *acting in accordance with a rule* is significant from the point of view of trying to understand decision making, as organisms that follow rules approach decision making in a very different way from ones that rely on a table of

[12] Note that this still allows us to characterize some of a drive-based organism's behaviors as *wrong*, even though it might be disposed to do all of them. For example, we could follow Millikan (1990) and say that a male hoverfly that darts at a bird and then gets eaten is not doing what it is (biologically) supposed to do – even though it might be reacting perfectly in line with its decision making mechanisms, and even though these mechanisms might have been selected for. This is because the reason these mechanisms were selected for (the 'distal rule' according to Millikan, 1990) was not to get the male hoverfly to dart at and get eaten by birds, but to intercept female hoverflies. Importantly, all of this is true whether the male hoverfly's decision making mechanisms are based on an extensive table of state of the world-action connections (with rows such as 'image of a small-sized object that is moving with angular velocity 100 degrees across the retina'' – 'go the spot that is 170 degrees of the image across the retina', and so on) or on an inner representation that says 'the thing to do is to accelerate towards the spot that is 180 degrees minus 1/10 of the angular vector velocity of the image of the fast moving small-sized object across the retina'. In other words: the distinction between *following a rule* and *acting in accordance with a rule* should not be conflated with the distinction between *following or acting in accordance with a specific behavioral rule* and *following or acting in accordance with any rule that is consistent with a certain behavioral pattern*. (Note also that neither of these should be conflated with the distinction between *following or acting in accordance with a rule* and *not acting at all*. A male hoverfly that is blown by the wind so as to intercept a female can be seen not to be acting at all: see e.g. Millikan, 1990, p. 333).

state of the world-action connections. In particular, rule following organisms engage in a form of practical reasoning: in making a decision, they *apply* a rule to a given situation – that is, they combine what they take the world to be like with what they think an appropriate way of dealing with the world is.[13] This contrasts with organisms that are drive-based, in that the latter merely *react* to the state of the world – they do not really engage in any kind of practical reasoning at all. Because of this, the distinction between following a rule and acting in accordance with a rule indeed is a useful way to mark desires from drives: it encapsulates the difference between practical reasoning and mere practical 'reacting' (see also Millikan, 2002).[14]

In short: the main issue to be assessed here is the question of why an organism would consult an explicit tokening of what it ought to do, when it could also have its actions simply be driven by behavioral dispositions that lack content. To answer this question, Sterelny argues that, in environments where behavioral flexibility is called for, desire-based organisms have several advantages over drive-based ones – advantages that can be seen as making the evolution of desires plausible. He specifically names four such advantages (see Sterelny, 2003, pp. 92-94):

1. *Desire-based organisms have an easier time making adaptive decisions when the range of behavioral options open to them is large.* Drive-based organisms run into problems when their decision problems get complex. If the only behavioral outcomes open to an organism are 'fight' or 'flight', drives might be adequate; however, if there are *n* different types of fighting behaviors,

[13] Note that, as made clear earlier (and see also note 23) what it means to take the world to be a certain way can be left open: it could involve a belief, but it might also just involve the excitation of more basic input systems (e.g. certain perceptions or tactile sensations).

[14] In what follows, I do not place any a priori constraints on the complexity or sophistication of the desires that an organism can entertain. Of course, this is consistent with the claim that organisms differ significantly in this respect (e.g. due to the presence of a natural language), but this deserves its own discussion (see e.g. Gallistel, 1990; Millikan, 1995, 2002; Carruthers, 2006; Bermudez, 2007).

and *m* different types of fleeing behaviors, it becomes much harder to see how a drive-based organism can reliably act adaptively.

2. *Desire-based organisms do not need to rely on a vast number of motivational states*. A sufficiently complex drive-based organism would need an astronomically large number of drives in order to be able to behave adaptively. However, such a large number of drives is biologically implausible.[15]

3. *Desire-based organisms do not depend on unreliable mechanisms for adjudicating among their motivational states.* Coordination among the many different active drives of a drive-based organism can get difficult. In particular, according to one of the major mechanisms for achieving this coordination – the 'winner take all' principle – the most 'urgent' drive is given complete control over the organism's behavior; however, this kind of control mechanism will fail when it is necessary to take the urgings of other drives into account in order to behave adaptively. Desire-based organisms can avoid these sorts of problems.

4. *Desire-based organisms are able to cope more quickly with changes in their needs.* Drive-based organisms can only change their motivational structures in evolutionary time, whereas desire-based organisms can *learn* what is good for them. This, though, will put the former at a disadvantage relative to the latter – they are less quick at adapting to their environment.[16]

[15] Sterelny attributes this point to Dickinson: see Sterelny (2003, p. 93).
[16] This argument bears some similarities to Robson (2001) and Samuelson (2001); see also note 3 above.

However, when considering this positive part of Sterelny's account in more detail, it becomes clear that it cannot be considered fully plausible. This is so for two reasons: on the one hand, Sterelny does not make sufficiently clear why desires can avoid the problems he alleges to exist for drives; on the other hand, he does not make sufficiently clear why drives do have these problems in the first place. To see this, reconsider Sterelny's four alleged advantages of desires over drives.

The main problem with Sterelny's alleged advantage 1 is that it is not spelled out in any kind of detail. It is just not clear what, exactly, the problem with 'complex decision points' (Sterelny, 2003, pp. 92-93; see also pp. 94-95) is meant to be for drives, and so it is not clear how desires could solve this problem. Why is it easier to figure out which actions one is to engage in if one can consult what one is to do directly than if one cannot? Why does the mere fact that one can consider explicit tokenings of what one is to do help in deciding among a large set of behavioral options? As will become clearer below, I think Sterelny is gesturing at an important issue here; however, as it stands, this gesture cannot be considered a plausible argument. Much more work is needed to spell out this point.

When it comes to advantage 2, the problems are twofold. On the one hand, it is not at all clear why a large number of drives is biologically implausible – or, indeed, how many drives would be necessary to reach this number. In particular, it is hard to see how one is to determine or justify an absolute number of drives that is 'too much' – in fact, it is not even clear what the right order of magnitude is here. On the other hand and more importantly, it is not clear why the situation is meant to be better for desire-based organisms: after all, they might also need a lot of desires in

order to be able to behave flexibly. Why is it more plausible to have many desires than to have many drives?[17]

Advantage 3 suffers from the same sort of dual problem. On the one hand, it is not clear why a drive-based organism *must* rely on a 'winner-take-all' mechanism to make decisions – after all, there is nothing intrinsic to drive-based organisms that makes this mechanism a necessity. For example, such an organism could also use a mechanism that gives different drives different weights in determining an action (a sort of 'vector addition' model of action determination). On the other hand, it is not clear why the same problem might not arise for desire-based organisms as well; after all, these too have to find a way of coordinating their many different desires. As a matter of fact, some of the most prominent accounts of how this coordination works are based on precisely the kind of 'winner-take-all' principle that Sterelny thinks will often break down (see e.g. Carruthers 2006; Selfridge & Neisser, 1960).[18] Given this, it is not clear why desire-based organisms are meant to be better off here than drive-based ones.[19]

Finally, the same applies to advantage 4. In the first place, it does not seem to be true that drive-based organisms need to be slow in changing their motivational states: in fact, new drives often are acquired very quickly and easily by many organisms (see e.g. Sherman et al., 1997, pp. 79-81; Mackintosh, 1983). On the other hand, it is not obvious that desire-based organisms *need* to be much faster at acquiring new motivational states: in fact, it is at least conceivable that such

[17] It could be that Sterelny here (and in relation to the first alleged benefit of desires) reasons as follows: a desire-based organism could just rely on a few ultimate desires (in combination with its beliefs) – i.e. it could just have the desire to fight and the desire to flee, and then use its beliefs (say) to determine the appropriate behavior for the case at hand (see Sterelny, 2003, pp. 87-90). If so, then his account would come closer to mine here; again, though, a more detailed picture of the kind of reasoning the organism engages in – and of the benefits this might bring to it – is necessary to make this account plausible.

[18] Even some neuroscientific accounts of decision making employ a principle of this sort: see e.g. Glimcher et al. (2005).

[19] Sterelny (in personal conversation) has claimed that the fact that desires have structured content can also be used to answer this question (see also note 17). However, it is not clear why this would be so: if I want that p (and believe that A is the best way to achieve p) and I want that q (and believe that B is the best way to achieve q), it is not clear how appealing to the details of what p and q (or A and B) are can help me decide what to do.

organisms acquire most of their (fundamental) desires innately, and that they only change them over evolutionary time.[20] At any rate, there seems to be nothing inherent in content-bearing mental states that makes this possibility more or less likely. Again, therefore, it is not clear what the relative benefits of desire-based organisms are meant to be in this respect.[21]

Overall, therefore, it becomes clear that we have not been given a fully compelling account of the evolution of desires. In particular, it is not clear that suitably sophisticated drive-based organisms could not find ways of (a) appropriately adjudicating among the different ways of satisfying their motivational states, (b) handling many different motivational states, (c) coordinating their motivational states, and (d) changing their motivational states quickly. Moreover, it is not clear why desire-based organisms should be any better at (a)-(d). Hence, as things stands, we still need an account of the evolution of desire. I present the outlines of such an account in the next section.

III. Drives, Desires, and Rules

I spell out my account of the evolution of desires in two steps: firstly, I present the details of the account, and secondly, I consider an objection that might be raised against it.

1. The Benefits of Rules

To see the structure of my account most easily, it is best to begin with an example.[22] Assume a purely drive-based organism finds itself in a mountainous region without water, but in need of

[20] This is less plausible for instrumental desires. However, as Sterelny (2003, pp. 87-90) himself notes, it is not so clear that we want to rely on these in this context (see also note 17 above). See also Sober & Wilson (1998) and Stich (2007) for more on this point.
[21] This is also important in the context of economic arguments for the evolution of preferences – see also note 3.
[22] Note that this example is meant to be very stylized – it is used as an expository device, and not meant to provide a compelling model of some actual case.

drink. Assume further that: (a) there are several sources of water in the vicinity, some higher up in the mountains than others; (b) the water sources are not obviously recognizable, and thus need to be searched for; (c) walking uphill is very costly, and (d) the mountain is rugged, so that not every water source is easily accessible from everywhere on the mountain. Given this, to act appropriately, the organism needs to connect a different action to all the positions on the mountain at which it might be: depending on where exactly it is located, it is adaptive to either stay and search for a water source near its current location, or to walk to a location further down the mountain, and search there.[23] Hence, in order to decide what to do, this drive-based organism needs to search through a long list of state of the world-action connections to identify the entry that corresponds to its (presumed) actual position, and act in line with that entry (see figure 1).[24]

[figure 1]

By contrast, if the organism has desires – so that it can consult an explicit tokening of the principle that drives its behavior – it no longer needs to map highly specific behavioral patterns with every state of the world it can distinguish, but can *calculate* what to do. So, for example, if the above organism acquires the ability to explicitly consult the principle behind its actions (to find water in the most efficient manner, taking into account the features of its environment) it can just use its (presumed) position as an input into the relevant function that determines the best

[23] Note that, in line with the fact (noted earlier) that I here want to avoid committing to whether the relevant organism already has beliefs, I shall here leave open the exact nature of the manner in which the organism detects the state of the world. In particular, this could be based on beliefs or on more basic input states; in either case, the point is just that the organism's take on what the state of the world is directly triggers certain behavioral patterns.

[24] Note that this is a point about the structure of the decision making mechanism, not its actual implementation. For example, the organism could use a form of content-addressable memory (or some other parallel search procedure) to help it find the appropriate line in the look-up table. This, though, would not affect the point made here that decision making on this model consists of dealing with a large 'look-up table' of state of the world-action connections. I thank Colin Allen for some useful remarks about these issues.

thing to do. Put differently, the ability to consult the content of a desire implies that the organism no longer needs to remember every instance of the behavioral rule it follows, but can just remember the rule, and apply it to the particular case it is in: effectively, these organisms can replace the entire table in figure 1 with the rule: *the (pro tanto) thing to do is to walk to location* $\sqrt{x}$ *(where x is the organism's current location) – rounded down to nearest integer if not itself an integer – and initiate search for water source there*.[25] Figure 2 presents an example of the kind of reasoning that this would lead to:

[figure 2]

This ability to avoid a large look-up table of state of the world-action connections matters, as it provides a reason for the evolution of desires. This reason turns on the fact that desire-based organisms can be more *cognitively efficient* – in a specific sense – than their drive-based competitors. In particular, the reliance on a desire – i.e. the consultation of an explicit behavioral rule – can reduce the memory requirements of the relevant organisms when compared to the drive-based alternative that merely leads to action in accordance with this rule. These memory savings stem from the fact that, instead of having to store *n* different state of the world-action connections, desire-based organisms just need to store the relevant behavioral function. In turn, this is important for two reasons.[26]

[25] This rule is qualified by a 'pro tanto' clause, as there might several different rules that are applicable to the situation at hand. In that case, the organism has to find a way of weighing up these different rules against each other. I shall not discuss this problem further, though, as it does not bear on the question at stake here: as noted above, this issue arises equally for desire-based and drive-based organisms.

[26] Samuelson (2001) also sees the cognitive efficiency of cognitive architectures as an important determinant of their evolution; however, his argument addresses a slightly different issue from the one relevant here (see also note 3 above).

Firstly, it plausibly leads to savings in the organism's *energetic* resources: assuming that the neural facts track the cognitive facts here – which is not unreasonable (see e.g. Ellis & Morgan, 1999) – the fewer state of the world-action connections an organism has to store, the less extensive the relevant neural network has to be that underlies its memory system. In turn, with a smaller neural network, the organism needs to build and maintain fewer neural pathways, and thus saves energy. These energetic savings can then be put to use in other places. Note that, for this to be true, it must be the case that storing desires does not require more memory – or memory that is of a different (e.g. declarative) and more energetically costly kind – compared to storing state of the world-action connections. Given the current state of knowledge concerning the different kinds of memory and their uses, this is not implausible (see e.g. Eichenbaum & Fortis, 2009).[27]

Secondly, the memory gains that come from the avoidance of a large look-up table of state of the world-action connections are beneficial for their own sake. For example, the organism can now keep track of more of the relationships among the individuals of its group, or learn to make finer distinctions in the categorization of its non-social environment. In this way, desire-based organisms come to have what is effectively a larger memory store than their drive-based competitors (assuming, again quite reasonably, that storing rules is not greatly more cognitively demanding than storing state of the world-action connections).

Now, it is important to note that these cognitive benefits of rule-based organisms may come with an important downside: these organisms plausibly have to pay higher costs in *calculating* what the appropriate response to their environment is. In particular, it may well be true that using behavioral rules to calculate the appropriate response to one's environmental situation can take a

[27] To the extent that it is questioned, though, it will reduce the efficiency gains provided by desire-based cognitive architectures – see also note 28.

longer time and require larger amounts of cognitive resources like concentration and attention than merely looking up this response in an appropriate table. Note, though, that the issues here are not entirely straightforward, in that drive-based organisms still have to deal with a large (possibly enormously so) database of state of the world-action connections, the handling of which is also likely to require much in terms of cognitive resources. Still, it may well be true that relying on behavioral rules can come with increased computational costs.[28]

However, this point should not be overemphasized. In particular, one must be careful here in not falling into the opposite extreme: while relying on desires may not be uniformly more cognitively efficient than relying on drives, it need not be uniformly less cognitively efficiently than relying on drives either. In other words, there is no reason to think that the costs of relying on desires *always* outweigh its benefits – the latter can be substantial, and the costs need not always be equally substantial. Where the balance of costs and benefits ends up depends on the details of the particular situation in question. In general, though, one would expect desires to be more efficient than drives when the relevant look-up table of behavioral responses is long – i.e. many states of the world are distinguished by the organism – and where the relevant rule that underlies these responses is relatively straightforward.

Clear cases that might exemplify these kinds of scenarios include some types of foraging and hunting situations: in these cases, the principle underlying the organism's behavior might be quite simple, but it might need to be applied to many different cases, and lead to many different consequences.[29] So, for example, if an organism hunts different kinds of prey, if these different

[28] As suggested in note 27, this conclusion would be strengthened if storing an explicit behavioral rule did turn out to be more energetically or cognitively costly than storing a state of the world-action connection.
[29] Interestingly, whether social interactions – a key element of Sterelny's (2003) account – favor desires or drives is not so clear. In particular, it is not clear how simple the relevant principles can be in this context, and how many different states of the world need to be distinguished. See e.g. Sterelny (2003), Boyd & Richerson (2005), and Sterelny (2012) for – somewhat contrasting – accounts of this.

kinds of prey call for different hunting strategies, and if these different hunting strategies are all versions of the same fairly simple principle, then desires might be cognitively more efficient than drives. Concretely, if the best way to kill the different kinds of prey is to tire them out, if there are many ways of tiring out prey, and if it is easy to figure out which of these is the right one in which situation – e.g. if it is true that the bigger the prey is, the more important it is to force it to exert a lot of effort in a short amount of time – then desires can help determine the appropriate hunting strategies more easily than drives.[30]

In this way, it becomes clear that desires can – though also need not – be more efficient than drives. These potential efficiency gains matter, as natural selection is known to favor efficiency where this is available (see e.g. Sockol et al., 2007; Roth-Nebelsick et al., 2001; Sober & Wilson, 1998; Stich, 2007). While organisms that are efficient in generating their behavior do not necessarily act more adaptively than their less efficient competitors *in any given context*, they are able to act more efficiently *overall* – they have more cognitive and energetic resources left than their less efficient competitors, which will prove useful to them in other contexts. In short: desire-based organisms can be more adaptive than drive-based ones, and that for reasons of cognitive efficiency.

At this point, it is important to pause and make clearer exactly what this conclusion establishes, and what not. The goal and upshot of the above discussion is a defense of the claim that there can be fitness differences between two traits: drive-based and desire-based minds. Now, the existence of fitness differences of this kind is, by itself, not sufficient to allow one to conclude either that the fitter trait must evolve or that this trait must evolve due to its being the fitter trait (see e.g. Sober & Wilson, 1998; Sober, 2000). There are two main reasons for this (see e.g. Sterelny & Griffiths, 1999; see also Dawkins, 1986): firstly, natural selection is not the only

[30] For more on the adaptive importance of different hunting strategies, see e.g. Sterelny (2012).

factor influencing the evolution of many traits (e.g. drift, migration, and inheritance biases are also important evolutionary determinants), and secondly, natural selection is constrained by various genetic, developmental, and environmental factors (i.e. what it can select and how it can select it). For a full evolutionary argument, these two limitations on adaptationist reasoning must therefore be carefully evaluated. This is especially important to keep in mind when the relevant traits are psychological: establishing the relative importance of and constraints on natural selection is often particularly difficult there (e.g. due to difficulties of inferring ancestral traits from the fossil record – though see also Mithen, 1990; Sterelny, 2012), and – maybe because of this – the necessary care in doing this has often been missing (see also Richardson, 2007; Buller, 2005).

However, despite these anti-adaptationist worries, the present account ought to be seen as having some epistemic value. This is so for two reasons. Firstly, natural selection is widely considered to be a sufficiently important evolutionary determinant to make the establishment of fitness differences between two traits *evidentially* relevant to the investigation of what evolves (see also Orzack & Sober, 1994). Particularly for complex traits (like mind designs), natural selection has often enough been shown to be an important causal factor so as to make it epistemically interesting to establish what it would entail about what should evolve – even in the absence of further knowledge about the importance of other evolutionary factors or constraints on natural selection (see also Dawkins, 1986; Godfrey-Smith, 2001; Sterelny, 2003). Secondly, nothing further than the establishment of *evidence* for a certain view of why and when desires should be expected to evolve is the aim of the present account: just like Sterelny (2003), my goal

here is merely to present *one plausible reason* for why some organisms might have switched to using desires. A full account of this evolution has to await further work on these issues.[31]

In all, therefore: the fact that desires can enable an organism to make decisions in a more efficient manner should be seen as a *prima facie* reason for why desires have evolved. Other things can interfere with this reason, but it should be seen as *a* reason nonetheless.

2. An Objection and a Reply

At this point, it is useful to discuss an objection that might be raised against this account. This objection concerns how desire- and drive-based organisms are to be empirically distinguished from each other. This is important, as it seems a prerequisite for the above evolutionary account of desire to be empirically testable: after all, in order to tell whether desires really evolved for the reasons laid out earlier, it would seem to be necessary to be able to say which organisms use drives, and which desires to make decisions – otherwise, it seems we have no data to which to apply this account. However (so the objection goes), it is not clear how this is to be done: as noted earlier, on my account, desire- and drive-based organism can give rise to identical behaviors. So how is one to tell how a specific organism's decision making architecture is organized?

In response, it should be noted that, while it may be true that empirically distinguishing desires from drives is often difficult, this does not mean that it is – and must always remain – impossible. What it is necessary is quite simply more data on human and non-human decision

[31] Put differently: fitness differences are *evidence* for the evolution of a given trait, but, by themselves, they should not be seen to force us to *accept* the hypothesis that this trait has evolved. Note also that the discussion of the objection raised below can be seen to suggest that such a fuller account of the evolution of desires is not impossible to achieve: for example, phylogenetic comparisons between those organisms that are drive-based and those that are desire-based, together with information about the ecological settings in which they evolved, might make it possible to further spell out and test the present account. I thank an anonymous referee for useful discussion of this point.

making (see also Sterelny, 2003, chap. 6). The set of data that seems especially important in this context concerns the kinds of decision making errors that an organism makes.[32]

Specifically, it seems reasonable to expect that the decision making errors of desire-based organisms can differ quite systematically from those of drive-based ones: the former will typically consist in incorrect applications of a rule, whereas the latter will typically consist in an organism selecting the wrong entry in its table of drives, or in its connecting a state of the world to the wrong kind of behavior. These two may be dissociable, in that the former can lead to completely new kinds of behaviors (i.e. behaviors which the organism would not normally do at all), whereas the latter will typically involve organism-typical behaviors that are just produced in the wrong circumstances. This distinction may be open to experimental and field-based investigation, thus giving us a means to distinguish drive-based and desire-based organisms.

Note that the point here is not that the two cognitive architectures *always* make different kinds of errors (indeed, it is to be expected that they do so only quite rarely), or that one typically makes *more* errors than the other, or that one typically makes more *costly* errors than the other. The point is just that they can, sometimes, make different kinds of errors. How often the errors of the two architectures differ, which architecture ends up being more error prone, and the typical errors of which architecture end up being more costly depends on the details of the case. For this reason, the different errors profiles cannot be used to argue in favor of one or the other of the two

[32] Another set of data that is often thought to be useful in distinguishing representational from non-representational cognitive architectures concerns patterns of functional degradation. In particular, it is frequently thought that cognitive architectures based on non-content-bearing mental states (like drive-based ones) are more likely to exhibit graceful degradation in function when compared to cognitive architectures based on content-bearing mental states (like desire-based ones) (see e.g. Rumelhart, McClelland, et al., 1986). However, there are two reasons to think that, in general, this may not be a particularly useful way of distinguishing the two kinds of cognitive architectures after all. Firstly, it is plausible that there are backup systems that cushion the degradation in the desire-based case (see also Sober & Wilson, 1998; Schulz, 2011) – for example, there might be a secondary rule that the organism can rely on, in case the primary one is inaccessibly or faulty. Secondly, some kinds of damage to drive-based architectures can lead to discontinuous drops in functioning – for example, those that completely dissociate the two columns of the table of state of the world-action connections, or those that prevent the organism from accessing this table at all. For this reason, differences in functional degradation patterns do not seem to be particularly useful as a diagnostic tool in this context.

cognitive architectures generally being more reliable than the other. However, this does not alter the point that the kinds of errors an organism makes are a useful *diagnostic tool* for assessing whether it is desire-based or drive-based.

In short: given enough data (of this and other kinds), the models of cognition that are found most useful in accounting for this data should provide a good guide as to which animals use drives and which use desires to make decisions (see also Allen, forthcoming). In turn, this data could then be used as part of phylogenetic comparative studies to come to a fuller picture of the evolution of desires – and, with this, to develop a better sense of the empirical merits of the present account. Note also that this account (as well as that of Sterelny, 2003) retains its interest even before this empirical work is far advanced: the goal of these accounts is to determine plausible *prima facie* reasons for why desire-based minds might have evolved. As made clearer below, this kind of account is interesting even before we are fully certain about which organisms use which model of the mind to make decisions (see also Millikan, 2002).

IV. The Biological Importance of Desires: An Implication

One of the key implications of this account of the evolution of desires concerns the importance of desire-like representations in an organism's cognitive life. To see this, begin by recalling that, on my account, relying on desires can – though need not – make it cognitively and energetically *easier* for an organism to make decisions: it allows it to avoid having to store many individual state of the world-action connections and thus helps it react more efficiently to a wide range of situations.

This matters, as precisely the opposite would be expected on many recent accounts of the nature of mental representation: a number of authors have come to argue that relying on mental

representations when making decisions is an avoidable luxury – the 'world is its own best model' and does not need to be re-presented in an organism's mind (see e.g. Clark, 2008; Rowlands, 1999; van Gelder, 1996; see also Shapiro, 2010, 2011). In particular, the view defended by these authors is that, in many circumstances, relying on mental representations just adds unnecessary cognitive labor to an organism – the environment contains enough cues to make adaptive action possible without needing to be supplemented by representational cognition (see e.g. Rowlands, 1999; Shapiro, 2010).

What my account adds to this debate is the claim that this sort of anti-representationalist stance should be neither overemphasized nor set aside. In particular, my account shows that, on the one hand, when looked at from an evolutionary point of view, mental representations should in fact often be expected to be the basis of an organism's decision making apparatus – and that precisely because they can *streamline* the way this organism determines how to best interact with its environment. Contrary to the anti-representationalist argument, then, the benefits that come from relying on desires – as representational motivational states – lie precisely in the fact that they make decision making *more*, not less, efficient.

On the other hand, though, my account also shows that the anti-representationalists are onto something. Since relying on desires is not universally adaptive – indeed, on my account, there are cases where relying on desires would be *maladaptive* – there are many cases where we should *not* expect (motivational) representations to underlie an organism's decision making apparatus.[33] This point is further strengthened by noting that the behavioral abilities of drive-based organisms should not be understated – in particular, these organisms should be seen to be

[33] This conclusion also gains support from the facts that natural selection is not the only factor that drives the evolution of organismic traits (as noted earlier), and that the evolution of desires from drives might well involve a costly transition period.

able to do everything that their representationalist competitors can.[34] For these reasons, the anti-representationalist stance should also be seen to get at something fundamentally right in this context.

In this way, the present account can help to balance the debate surrounding the importance of representations in an organism's cognitive life, and brings out what is plausible about both sides. Importantly, moreover, it can make it a bit easier to settle this debate, as we now have some idea of *when* we would organisms to be desire-based, and when drive-based. In particular, as noted earlier, to the extent that these organisms evolved in conditions that required keeping track of many different states of the world, and where it was possible to react to these states with actions that it was relatively easy to determine by consulting a behavioral rule, we should expect them to be desire-based – and the opposite. While not settling this debate, therefore, my account can help move it a bit closer to its resolution.

V. Conclusion

I have tried to show that a *prima facie* reason for why desires have evolved can be seen to lie in their enabling an organism's decision making systems to become more efficient in a specific manner: they do not require the organism to store a large set of behavioral patterns, but allow it to decide what to do by calculating the appropriate response to its environment. This stands in contrast to Sterelny's (2003) account, which does not pinpoint a narrowly circumscribed benefit of desires, but rather presents various reasons for why desires might be useful in environments rewarding behavioral complexity. I also hope to have shown that this account has important

[34] This also explains why we should not expect all the actions of one organism to be desire-driven – in particular, it is to be expected that an organism will maintain some drives (e.g. if the selection pressure to make efficient decisions was not equally large in all decision contexts). In fact, the present account lends itself to be spelled out in terms of dual process models like that of Haidt (2001, p. 818): these models combine an older 'associative' system and a newer 'rule-driven' system to make sense of the hybrid nature of many human psychological capacities.

implications concerning the value of mental representations in an organism's mental life. In this way, it is hoped that the present theory goes at least some ways towards improving our understanding of the belief / desire model of the mind and its evolutionary presuppositions.

Bibliography

Allen, C.: 2004, Is Anyone a Cognitive Ethologist?, Biology and Philosophy 19: 589-607.

Allen, C.: forthcoming, Models, Mechanisms, and Animal Minds.

Allen, C., and Bekoff, M.: 1997, Species of Mind, MIT Press, Cambridge, MA.

Anscombe, E.: 1957, Intention, Cambridge University Press, Cambridge.

Bekoff, M., Allen, C., and Burghardt, G. (eds.): 2002, The Cognitive Animal, MIT Press, Cambridge, MA.

Bermudez, J.: 2007, Negation, Contrariety, and Practical Reasoning: Comments on Millikan's 'Varieties of Meaning', Philosophy and Phenomenological Research 75: 663-669.

Boghossian, P.: 1989, The Rule-Following Considerations, Mind 98: 507-549.

Boyd. R., and Richerson, P. (2005).: The Origin and Evolution of Cultures. Oxford: Oxford University Press.

Buller, D.: 2005, Adapting Minds, MIT Press, Cambridge, MA.

Carruthers, P.: 2006, The Architecture of Mind, Cambridge University Press, Cambridge.

Cheng, K.: 2011, A New Look at the Problem of Rule Following: A Generic Perspective, Philosophical Studies 155: 1-21.

Clark, A.: 1991, In Defense of Explicit Rules, In W. Ramsey, S. Stich, and D. Rumelhart (eds.), Philosophy and Connectionist Theory, Lawrence Erlbaum, Hilldale, pp. 115-128.

Clark, A.: 1992, The Presence of a Symbol, Connection Science 4: 193-205.

Clark, A.: 2008, Supersizing the Mind, Oxford University Press, Oxford.

Davidson, D.: 1982, Rational Animals, Dialectica 36: 318–27.

Davidson, D.: 1987, Knowing One's Own Mind, Proceedings and Addresses of the American Philosophical Association, 60: 441-58.

Dawkins, R.: 1986, The Blind Watchmaker, Norton, New York.

Dickinson, A., and Balleine, B.: 2000, Causal Cognition and Goal-Directed Action, In C. Heyes and L. Huber (eds.), The Evolution of Cognition, MIT Press, Cambridge, MA, pp. 185-204.

Dretske, F.: 1981, Knowledge and the Flow of Information, MIT Press, Cambridge, MA.

Dretske, F.: 1988, Explaining Behavior, MIT Press, Cambridge, MA.

Eichenbaum, H., & Fortin, N.: 2009, The Neurobiology of Memory Based Predictions. Philosophical Transactions of the Royal Society B, 364: 1183-1191.

Ellis, D. & Morgan, N.: 1999, Size Matters: An Empirical Study of Neural Network Training for Large Vocabulary Continuous Speech Recognition, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1013 – 1016.

Fodor, J.: 1990, A Theory of Content, In A Theory of Content and Other Essays, MIT Press, Cambridge, MA, pp. 51-135.

Gallistel, R.: 1990: The Organization of Learning, MIT Press, Cambridge.

Glimcher, P., Dorris, M., and Bayer, H.: 2005, Physiological Utility Theory and the Neuroeconomics of Choice, Games and Economic Behavior 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 E. Sober and S. Orzack (eds.), Adaptationism and Optimality, Cambridge University Press, Cambridge, pp. 335-357.

Guth, W.: 1995, An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives, International Journal of Game Theory 24: 323-344.

Harre, R.: 2002, Social Reality and the Myth of Social Structure, European Journal of Social Theory 5: 111-123.

Haidt, J.: 2001, The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment, Psychological Review 108: 814-834.

Hausman, D.: 2012, Preference, Value, Choice, and Welfare, Cambridge University Press, Cambridge.

Heyes, C. and Huber, L. (eds.): 2000, The Evolution of Cognition, MIT Press, Cambridge, MA.

Kirsh, D.: 1996: Today the Earwig, Tomorrow Man? In M. Boden (ed.), The Philosophy of Artificial Life, Oxford: Oxford University Press, pp. 237-261.

Kripke, S.: 1982, Wittgenstein on Rules and Private Language, Harvard University Press, Cambridge, MA.

Lewis, D.: 1988, Desire as Belief, Mind 97: 323-332.

Lewis, D.: 1996, Desire as Belief II, Mind 105: 303-313.

Mackintosh, N. J.: 1983, Conditioning and Associative Learning, Clarendon Press, Oxford.

McFarland, D.: 1996, Animals as Cost-Based Robots. In M. Boden (ed.), The Philosophy of Artificial Life, Oxford: Oxford University Press, pp. 179-207.

Mele, A.: 1987, Are Intentions Self-Referential?, Philosophical Studies 52: 309-329.

Millikan, R.: 1984, Language, Thought, and Other Biological Categories, MIT Press, Cambridge, MA.

Millikan, R.: 1990, Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox, The Philosophical Review 99: 323-353.

Millikan, R.: 1995, Pushmi-Pullyu Representations, Philosophical Perspectives, 9: 185-200.

Millikan, R.: 1996, On Swampkinds, Mind and Language, 11: 70-130.

Millikan, R.: 2002, Varieties of Meaning, MIT Press, Cambridge, MA.

Mithen, S.: 1990, Thoughtful Foragers, Cambridge University Press, Cambridge.

Neander, K., 2012: Toward an Informational Teleosemantics. In J. Kingsbury and D. Ryder (ed.), Millikan and Her Critics, Oxford: Wiley Blackwell, pp. 21-41.

Nichols, S., & Stich, S.: 2003, Mindreading, Oxford University Press, Oxford.

Orzack, S., & Sober, E.: 1994: Optimality Models and the Test of Adaptationism, American Naturalist, 143: 361-380.

Papineau, D.: 1987, Representation and Reality, Blackwell, Oxford.

Papineau, D.: 2004, Friendly Thoughts on the Evolution of Cognition, Australasian Journal of Philosophy 82: 491-502.

Prinz, J.: 2002, Furnishing the Mind, MIT Press, Cambridge, MA.

Railton, P.: 2012, That Obscure Object, Desire, Proceedings and Addresses of the American Philosophical Association 86: 22-46.

Richardson, R.: 2007, Evolutionary Psychology as Maladapted Psychology, MIT Press, Cambridge, MA.

Robson, A.: 2001, Why Would Nature Give Individuals Utility Functions?, The Journal of Political Economy 109: 900-914.

Robson, A., and Samuelson, L.: 2008, The Evolutionary Foundations of Preferences, In J. Benhabib, A. Bisin, and M. Jackson (eds.), The Social Economics Handbook, Elsevier Press, Amsterdam.

Roth-Nebelsick, A., Uhl, D., Mosbrugger, V., and Kerp, H.: 2001, Evolution and Function of Leaf Venation Architecture: A Review, Annals of Botany 87: 553-566.

Rowlands, M.: 1999, The Body in Mind: Understanding Cognitive Processes, Cambridge University Press, Cambridge.

Rumelhart, D., McClelland, J., et al.: 1986, Parallel Distributed Processing, Vol. I, MIT Press, Cambridge, MA.

Samuelson, L.: 2001, Analogies, Adaptation, and Anomalies, Journal of Economic Theory 97: 320-366.

Schroeder, T.: 2004, Three Faces of Desire, Oxford University Press, Oxford.

Schroeder, T.: 2009, Desire, In Edward N. Zalta (ed.), Stanford Encyclopedia of Philosophy, Winter 2009 Edition, URL = <http://plato.stanford.edu/archives/win2009/entries/desire/>.

Schulz, A.: 2009. Condorcet and Communitarianism: Boghossian's Fallacious Inference, Synthese, 166: 55-68.

Schulz, A.: 2011: Simulation, Simplicity, and Selection, Philosophical Studies, 152: 271-285.

Selfridge, O., and Neisser, U.: 1960, Pattern Recognition by Machine, Scientific American, 203: 60-68.

Shapiro, L.: 2010, James Bond and the Barking Dog, Philosophy of Science 77: 400–418.

Shapiro, L.: 2011, Embodied Cognition, Routledge, Oxon.

Sherman, P., Reeve, H., Pfennig, D.: 1997, Recognition Systems, In J. Krebs & N. Davies (eds.), Behavioural Ecology: An Evolutionary Perspective, Blackwell, Oxford, pp. 69-96.

Skyrms, B.: 2010. Signals: Evolution, Learning, and Information. Oxford: Oxford University Press.

Smith, M.: 1987, The Humean Theory of Motivation, Mind 96: 36-61.

Sober, E.: 1994, The Adaptive Advantage of Learning and A Priori Prejudice, In: From A Biological Point of View, Cambridge University Press, Cambridge, pp. 50-70.

Sober, E.: 2000, Philosophy of Biology, Second Edition, Westview, Boulder.

Sober, E. & Wilson, D. S.: 1998, Unto Others: The Evolution and Psychology of Unselfish Behavior, Harvard University Press, Cambridge, MA.

Sockol, M, Raichlen, D., and Pontzer, H.: 2007: Chimpanzee Locomotor Energetics and the Origin of Human Bipedalism, Proceedings of the National Academy of Science, 104: 12265-12269.

Spier, E, & McFarland, D.: 1998, Learning to Do Without Cognition, In R. Pfeiffer, B. Blumenberg, J. A. Meyer, and S. Wilson (eds.), From Animals to Animats 5, Cambridge, MA: MIT Press, pp. 38-47.

Stampe, D.: 1986, Defining Desire, In J. Marks (ed.), The Ways of Desire, Precedent Publishing, Chicago, pp. 149-174.

Sterelny, K.: 1991, The Representational Theory of Mind, Blackwell, Oxford.

Sterelny, K.: 1999, Situated Agency and the Descent of Desire, In V. Hardcastle (ed.), Where Biology Meets Psychology, Cambridge, MA: MIT Press, pp. 203-220.

Sterelny, K.: 2001, The Evolution of Agency, In The Evolution of Agency and Other Essays, Cambridge: Cambridge University Press, pp. 260-287.

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

Sterelny, K.: 2012. The Evolved Apprentice. Cambridge, MA: MIT Press.

Sterelny, K & Griffiths, P.: 1999, Sex and Death, University of Chicago Press, Chicago.

Stich, S.: 2004, Some Questions from the Not So Hostile World, Australasian Journal of Philosophy 82: 491-498.

Stich, S.: 2007, Evolution, Altruism and Cognitive Architecture: A Critique of Sober and Wilson's Argument for Psychological Altruism, Biology and Philosophy 22: 267-281.

van Gelder, T.: 1996, Dynamics and Cognition, In J. Haugeland (ed.), Mind Design II, MIT Press, Cambridge, MA, pp. 421-450.

Figures

The organism's (presumed) current location	**The location at which to search for a water source**
0	0
1-3	1
4-8	2
9-15	3
16-24	4
25-35	5
36-48	6
49-63	7
64-80	8
81-99	9
100-120	10
121-143	11
144-Summit	12

[Figure 1: Representation of a Drive-Based Cognitive Architecture]

(Pro Tanto) Rule: Walk to location $\sqrt{x}$ (where x is the organism's current location) – rounded down to nearest integer if not itself an integer – and initiate search for water source there.
Presumed location: 3.
(Pro Tanto) Action: Walk to location 1 (=$\sqrt{3}$ rounded down); initiate search for water source.

[Figure 2: Representation of a Desire-Based Cognitive Architecture]