THE CIRCULARITY OF EVOLUTIONARY RELIABILISM

James Sage
Department of Philosophy
University of Wisconsin-Stevens Point
jsage@uwsp.edu

I. INTRODUCTION

This paper addresses an exchange between Alvin Plantinga and William Ramsey regarding the reliability of human cognitive faculties, given that they result from natural selection. This issue is important for philosophers who want to offer an account of epistemology, while at the same time maintaining some sort of naturalism. In other words, many philosophers hold both of the following claims:

(N)   human cognitive faculties result from natural selection.

(R)   human cognitive faculties reliably produce true beliefs.

Plantinga (1993, 2002) argues that accepting (N) provides good reason to doubt (R). But if there is good reason to doubt (R), then there is good reason to doubt any belief produced by human faculties, including the belief in (N) itself. And so belief in (N) is epistemically self-defeating. Ramsey (2002) attempts to rescue (N) from Plantinga's attack. His attempt is based on a position he calls "evolutionary reliabilism."

According to Ramsey, evolutionary reliabilism has two goals. The first goal is to show that (N) is not epistemically self-defeating. This is accomplished by constraining on the range of relevant beliefs such that the truth-ratio of this range of beliefs is in fact fairly high. And so,

when human cognitive faculties are epistemically appraised, we will find that they are indeed

reliable and the resulting beliefs are warranted.

The second goal of evolutionary reliabilism is to explain adaptive behavior in terms of

the truth of certain types of beliefs. Ramsey attempts to constrain the set of relevant beliefs the

truth of which best explains adaptive success.

I object to evolutionary reliabilism by arguing that it relies on a circular account of how

to appraise our cognitive faculties as well as a circular account of how to best explain adaptive

behavior in terms of truth. If I am correct, then Plantinga's argument against (N) is left

unscathed. In addition to this, I develop an objection to evolutionary reliabilism that further

strengthens Plantinga's anti-naturalism argument.

II. PLANTINGA'S ARGUMENT

I have reconstructed Plantinga's (1993) argument is as follows:

(1) Evolution selects for reproductive fitness and not necessarily for having true beliefs.
(2) There are numerous high-fitness designs for cognitive organisms.
(3) The portion of high-fitness designs according to which the organism has cognitive faculties
     that reliably produce true beliefs is either relatively low or inscrutable.
(4) Therefore, the conditional probability of (R) given (N) is either relatively low or inscrutable.
(5) The warrant of a belief depends in part on the reliability of the faculties that produce it.
(6) If (N) is true, then the warrant of all of our beliefs is either low or inscrutable.
(7) If (N) is true, then the warrant of our belief in (N) is either low or inscrutable.

I take it that (1) and (2) are uncontroversial. (4) follows from (2) and (3). (5) follows from

Plantinga's account of warrant (an account defended by Plantinga [1993] and accepted by any

proponent of evolutionary reliabilism, including Ramsey [2002]). (6) follows from (1) through (5). (7) follows from (6). So, the controversial move in the anti-naturalism argument is (3).

Plantinga argues for (3) in two ways. First, he cites examples where an organism enjoys a high degree of fitness despite having "deep misconceptions" about the world. Ramsey adequately rebuts this "deep misconceptions" argument, and so I will not consider it further.

The second way Plantinga supports (3) is to construct examples in which an organism with a high degree of fitness has mostly false beliefs, and this is achieved by modifying an organism's belief-desire pairs.

For example, consider Smith who, when confronted with a predator, is rewarded (i.e., gets to continue living) for behaving in certain adaptive ways (by fleeing or finding a place to hide). Now, Smith's adaptive behavior may be due to the following belief-desire pair: Smith desires to be eaten by the predator and believes that his chances of being eaten are better if he can find a hiding place (because predators, of course, enjoy a good challenge). Or, perhaps Smith desires to *approach* the predator, believing (falsely) that the predator is in fact a cuddly creature, but believes (falsely) that the best way to get closer to it is to run away from it (Plantinga, 1993: 226). Clearly, a large number of (false) belief-desire pairs can explain the adaptive behavior Smith exhibits. False beliefs, together with the right sort of desires, gets Smith out of harm's way.

Ramsey finds these examples implausible, and this is because of their piecemeal nature—it is hard to see how an organism, like Smith, who holds the belief "in order to get closer to something I should run away from it" will survive very long in the real world. Just imagine how hard it would be for Smith to find food or successfully reproduce while at the same time holding

such a belief. Ramsey maintains that what we need is a systematic account of how unreliable cognitive systems can nevertheless be adaptive.

Plantinga can support (3) by appealing to the evolutionary phenomenon known as the Garcia effect.[1] To illustrate this phenomenon consider the situation of Paul, who desires to avoid predators and has reasonable beliefs about them (i.e., he does *not* believe that the best way to approach something is to run away from it). However, in Paul's environment, there are two kinds of creatures that are perceptually indistinguishable to Paul. One kind of creature is a deadly predator, while the other is harmless. When Paul encounters either creature, he exhibits the same avoidance behavior—he hides. He also forms the belief with the content "the environment is dangerous" but this belief is usually false since harmless creatures vastly outnumber dangerous predators.

So, Paul has mostly false beliefs about predators in his environment but nonetheless behaves adaptively. In this way, the Garcia effect supports Plantinga's claim that an organism can be adaptive despite having cognitive faculties that generate mostly false beliefs. This further reaffirms the claim that natural selection cares about reproductive fitness, not necessarily about true beliefs. If a cognitive faculty happens to generate mostly true beliefs, then it is a happy accident from the point of view of evolution and fitness. This allows Plantinga to infer that:

(3) the portion of high-fitness designs according to which the organism has cognitive faculties that reliably produce true beliefs is either relatively low or inscrutable.

From (2) and (3) Plantinga concludes that:

(4) the conditional probability of (R) given (N) is either relatively low or inscrutable.

At first, it is tempting to conclude that if (4) is true, then we have good reason *not* to accept the conjunction of (N) and (R). By itself, however, (4) implies nothing about whether or not it is rational for us to accept both (N) and (R). For instance, the low probability of someone's being named Susan given that she is female does not make it rational for us to reject the conjunction that she is both female and named Susan. Independent evidence may show that she is both female and named Susan.[2]

Likewise, it is not rational to reject the conjunction of (R) and (N) on the basis of (4) alone if there is independent evidence in favor of both (R) and (N). For Plantinga, however, the only independent reasons for accepting (R) are theological reasons which are incompatible with (N). If Plantinga is correct, then for the proponent of (N), accepting (4) amounts to having good reason to doubt (R).

If all this is so, then we have reason to doubt any belief produced by cognitive faculties that result from natural selection. The reason for this is either (a) that any such belief is more likely to be false than true because the conditional probability of (R) given (N) is low, or (b) that the appropriate attitude to take regarding the truth of any such belief is one of agnosticism because the conditional probability of (R) given (N) is inscrutable. The final step of Plantinga's argument is to apply this result to (N) itself. From this we get:

(7) if (N) is true, then the warrant of our belief in (N) is either low or inscrutable.

The predicament is this: if the naturalist holds both (R) and (N), then naturalist must show that having true beliefs is itself adaptive. This is exactly what Ramsey attempts to do.

III. RAMSEY'S EVOLUTIONARY RELIABILISM

In this section, I will briefly present Ramsey's account, and then I will raise a number of objections in the next section. According to Ramsey,

> If all we want is an explanation of [Paul's] behavior, then we needn't appeal to the truth or falsehood of any of his beliefs. But if we want to know *why* [his] behavior proves *successful* … then it clearly matters that the behavior is generated by *true* beliefs. (Ramsey, 2002: 4-5, italics added)

In other words, the best explanation for why a cognitive organism's behavior is adaptive involves the claim that the beliefs that cause it to engage in that behavior are *true*. Although Ramsey acknowledges the initial difficulty presented by the Garcia effect, he insists that having reliably true beliefs *where it matters most* is what makes behaviors adaptive. Ramsey argues that the relevant range of Paul's beliefs is not simply all of his beliefs, regardless of their content. Rather, according to Ramsey, the content of an agent's beliefs "matters a great deal. It is the truth ratio of certain *types of* beliefs that matters" (Ramsey, 2002: 12).

For example, having true beliefs with the content "the square root of 25 is 5" would be less important than having true beliefs with the content "the environment is dangerous." Similarly, beliefs with the content "the environment is dangerous" would be less important than those with the content "the environment is safe." The reasoning is this: having false beliefs with the content "the environment is dangerous" is *less costly* than having false beliefs with the content "the environment is safe." A simple game theoretical matrix (see Table 1) can help us see that having false positives (believing that the environment is dangerous when it is not) is less costly than believing false negatives. For Paul, this amounts to employing a cautious belief-forming strategy—Paul tends to think that the environment is dangerous even when it is not.

[Table 1]

(what Paul believes)	There is a predator nearby	There is no predator nearby
that the environment is dangerous	+ a lot (duck in a hole & avoid being eaten)	– not much (duck in a hole)
that the environment is safe	– a lot (business as usual & get eaten)	+ not much (business as usual)

The first row of Table 1 depicts a cautious belief-forming strategy because it avoids the greatest cost—it has the best worst outcome. Having false beliefs with the content "the environment is safe" is disastrous—Paul gets eaten. Having false beliefs with the content "the environment is dangerous" is less costly—Paul ducks into a hole, and ducking into a hole is a minor disruption compared to getting eaten. False positives are tolerated; false negatives lead to Paul's early death.

According to Ramsey, the kind of organism rewarded might in fact employ a cautious strategy such as this. Paul's cautious strategy is adaptive *precisely because* he has the relevant kinds of *true* beliefs: he has true beliefs *when it matters most*, that is, when having false beliefs would be highly costly or maladaptive.

On Ramsey's account, as I understand it, we should sort Paul's beliefs into two kinds, those that are adaptively relevant, and those that are not. Adaptively relevant beliefs are those beliefs that are highly costly when false. It is the truth-ratio of these adaptively relevant beliefs

that we should use to appraise the reliability of the cognitive faculty in question. On Ramsey's account, an organism's set of adaptively relevant beliefs will have a high truth-ratio, and so the cognitive faculty responsible for generating these beliefs will be reliable.

Ramsey claims that the Garcia effect is not a difficulty at all. He claims that the Garcia effect helps us to understand how adaptive behavior is best explained by true beliefs: Paul would not be successful if he were not getting things right when it mattered most. It makes perfect sense that agents will have clusters of false beliefs. But because these false beliefs are relatively harmless, their presence is not a threat to the organism's adaptive success. The Garcia effect, therefore, is not a threat to Ramsey's evolutionary account explaining adaptive behavior in terms of truth.

Ramsey's evolutionary reliabilism has the following two goals. First, he argues that accepting (N) is *not* epistemically self-defeating. He argues that the conditional probability of (R) given (N) is in fact fairly high. When appropriately constrained, our beliefs will have a high truth-ratio, and so when we appraise human cognitive faculties, we find that they are indeed reliable.

Second, he argues that the best explanation for an organism's adaptive behavior is that it is caused by true beliefs with a certain type of content (the type of content wherein false negatives are highly maladaptive). This is an attempt to connect an organism's adaptive success with the truth of (a certain type of) its beliefs.

These two goals (the *epistemic appraisal* goal and the *explanation of adaptiveness* goal) are related in the following way. Natural selection operates on the basis of adaptiveness, so in order for a cognitive faculty to be naturally selected, the beliefs generated by it must contribute

to the our adaptive success. To show that the conditional probability of (R) given (N) is high is to explain adaptive success in terms of the truth of certain types of our beliefs.

IV. CIRCULARITY OBJECTION

When we epistemically appraise the reliability of an organism's cognitive faculties, Ramsey says that we should consider the truth-ratio of just *some* of its beliefs. Specifically, the relevant beliefs for epistemic appraisal are those with the type of content where having false beliefs is highly costly or maladaptive. So, when performing our appraisal we should consider just those beliefs the truth of which is adaptive. But, if we are looking at just those beliefs the truth of which is adaptive, and these beliefs are held by an adaptively successful organism, then, *a priori*, we can expect to find that the organism's set of beliefs has a high truth-ratio. And so it will always turn out (necessarily) that an adaptively successful organism's cognitive faculties are reliable.

This account, I claim, is *circular*. Ramsey holds that adaptively relevant beliefs are those beliefs the falsity of which is highly maladaptive. But to say that having false beliefs is maladaptive is to say that having true beliefs is adaptive. And so the adaptively relevant beliefs are just those beliefs the truth of which explains its success. The truth-ratio of this range of beliefs is high because the range of beliefs to be considered relevant is constrained to just those cases where truth is in fact adaptive. If I am right, then Ramsey's account is circular and so fails show that human cognitive faculties are reliable, given that they are produced by natural selection.

Circularity also infects Ramsey's attempt to show that the best explanation for adaptive behavior is having true beliefs. He needs to show that an organism's success is best explained by having true beliefs, but it cannot be the truth of an organism's beliefs *in general*. It has to be the truth of beliefs *with a certain type of content*. The relevant range of beliefs, then, must be constrained to include just those cases where having false beliefs is very costly. To claim that these false beliefs are costly is to claim that they are *not* adaptive, which is to claim that having true beliefs when it matters most *is* adaptive. In this way, Ramsey's account smuggles the notion of truth into the constraint itself, which spoils the whole point of trying to explain success in terms of truth.[3]

In other words, the criteria used to establish this range of adaptively relevant beliefs is determined in part by considerations of truth, and so the notions of truth and adaptiveness collapse. If truth is supposed to explain adaptiveness in any interesting way, then these notions must have enough "conceptual space" separating them. The problem is that Ramsey runs them together. At bottom, the range of beliefs whose truth values we should consider will be just those *where it matters most*. But *where it matters most* amounts to those cases where having false beliefs is maladaptive or, to put it bluntly, it amounts to just those cases where having true beliefs is in fact adaptive. But then it is no surprise that having true beliefs is what best explains the adaptive behavior of an adaptively successful organism—the range of relevant beliefs will be both true and adaptively successful because the notions of truth and success were used in determining the criteria for relevance in the first place.

If I am right about this circularity, then evolutionary reliabilism fails achieve its goals. Evolutionary reliabilism fails to give an account of how to appraise the reliability of our cognitive faculties. It also fails to give an account of how true beliefs are supposed best explain

adaptive behavior. The failure of Ramsey's evolutionary reliabilism leaves Plantinga's skeptical

argument against (N) unscathed.

V. AN ADDITIONAL DIFFICULTY

In this section, I argue that the proponent of evolutionary reliabilism needs to show that

(N) is warranted. In order for (N) to be warranted, however, the specific cognitive faculty

responsible for generating (N) must itself be reliable.

Recall that Ramsey's criteria for adaptively relevant beliefs is the high cost of having

false beliefs. Suppose reliable cognitive faculties include the "predator" belief faculty and the

"poisonous plant" belief faculty, to name just two. Is the "belief in Naturalism" faculty also

reliable? How is it disastrous to have a false belief with the content "human cognitive faculties

are the product of natural selection"? I do not see the disaster. A committed naturalist is not any

less fit simply in virtue of her belief in (N). Moreover, in terms of biological value, the content

of the belief in (N) is quite different from the content of Paul's predator beliefs. Hence, the

naturalist's adaptive success provides no reason to believe that the cognitive faculty responsible

for (N) is in fact reliable.

However, (N) may not be produced by a unique "belief in Naturalism" cognitive faculty.

Suppose cognitive faculty, K, generates both (N) and also a number of adaptively relevant

beliefs. Like Paul's "predator" belief faculty, K produces beliefs with two kinds of content—

those that are disastrous when false and those that are not. Suppose further that this set of

adaptively relevant beliefs generated by K has a high truth-ratio. So, even if we grant that K is

reliable, nothing follows regarding the probable truth of (N). While all beliefs produced by K

will share the same degree of warrant, we have no reason to think that the warrant enjoyed by (N) is any indication of its truth. After all, (N) was not part of the epistemic appraisal of K.

Regarding the warrant for the naturalist's belief in (N), it is a happy accident that K is reliable. It is a happy accident because K is reliable precisely because some *other* set of beliefs (those that are adaptively relevant) has a high truth-ratio. The warrant enjoyed by (N), therefore, is a by-product of a set of beliefs with radically different content, and suggests nothing about the truth of (N). If there is no reason to think that the truth of (N) is (or was) under selection pressure (i.e., adaptively relevant), then there is no reason to think that (N) is probably true. And if there is no reason to think (N) is probably true, then this fails to support the claim that the conditional probability of (R) given (N) is fairly high. All of this suggests that evolutionary reliabilism fails to rebut Plantinga's conclusion that belief in (N) is epistemically self-defeating.

VI. SOCIAL TRANSMISSION OF BELIEFS

Many beliefs are socially learned. The social division of cognitive labor for belief acquisition extends beyond individuals and beyond reproductive generations.[4] The cultural transmission of adaptive beliefs relieves evolution of the very difficult task of "hardwiring" each of us about modern dangers, such as touching hot stoves, playing on the freeway, or drinking antifreeze. Beliefs about these dangers are socially learned and confer an adaptive advantage to those who hold them. There is no way for natural selection to get these beliefs into our genes in any speedy way, even indirectly via specialized cognitive faculties. Cultural institutions serve as repositories for beliefs that contribute to our success. There is no need to postulate the existence of domain-specific, gene-linked cognitive faculties designed to generate these specific beliefs.

This alternative explanation locates the origin, persistence, and storage of such beliefs in human culture, not the human genome. What is required of evolution on this view is only that human cognition is sufficiently "plastic" so as to facilitate uptake of cultural information, such as the capacity to learn language and to pass along adaptive beliefs to others. Moreover, much of this is already present in our understanding of human cognition, linguistics, and human culture.

Interestingly, however, this cultural transmission account does not invoke the highly contested assumption that adaptiveness is best explained by true beliefs. Any belief, regardless of its truth value, can be stored in cultural institutions. The beliefs that are disastrous get weeded out; those that are adaptive stick around (for now). The truthfulness of our beliefs is a by-product of their adaptiveness, not a central feature of them.

VII. CONCLUSION

There are lots of ways to behave adaptively. Having true beliefs is certainly one such way. In a perfect world, having reasonable desires and true beliefs is, *cateris paribus*, adaptive. Unfortunately, we did not evolve in a perfect world, nor do we live in a perfect world now. In other words, *cateris* is not always *paribus*—we are surrounded by deception and camouflage and, in some cases, we are held captive by our own cognitive illusions. Falsehoods abound. The point is that there are a number of systematic ways for false beliefs to cause adaptive behavior. The Garcia effect is one such strategy. There are probably many other systematic strategies by which false beliefs lead to adaptive behavior, such as cheating, self-deception, and unwarranted optimism.

I have argued that Ramsey's version of evolutionary reliabilism is circular and thus fails solve the problem introduced by Plantinga's skeptical argument against (N). First, Ramsey fails to show that naturally selected cognitive faculties are reliable with respect to generating true beliefs. When we epistemically appraise our cognitive faculties, evolutionary reliabilism provides no reason to suppose that they are reliable, and so we have no reason to think that any of our beliefs are warranted. Second, Ramsey fails to show that the truth of our beliefs is what best explains our adaptive success. Based on the framework provided by evolutionary reliabilism, it is not at all clear that the truthfulness of beliefs is under any selection pressure whatsoever. In addition to these difficulties, I agued that the evolutionary reliabilist must take on the very difficulty task of showing that (N) is probably true. But because the content of (N) does not have any significant biological value (it is not costly when false), (N) is not under selection pressure. Hence, (N) is not likely to be true merely in virtue of being held by an adaptively successful organism.

It is not clear, however, that the epistemological naturalist has no other course of action. I suggested one such alternative account, a social transmission account, in order to explain human adaptiveness. This account is consistent with evolution and naturalism, it does not invoke theism, nor does it invoke the contentious assumption linking truth with adaptive behavior.

NOTES

[1] See, for example, Garcia, et al. (1972), Stich (1990: 61-63), and Stein (1996: 190-197).

[2] Johnny Cash's song "A Boy Named Sue" comes to mind, but this would count as independent evidence.

[3] For a similar charge of circularity, see Downes (2000) and Godfrey-Smith (1996: 174).

[4] See Downes (2000).

ACKNOWLEDGEMENTS

REFERENCES

Beilby, J., ed. (2002): *Naturalism Defeated: Essays on Plantinga's Evolutionary Argument Against Naturalism*, New York: Cornell University Press.

Dennett, D. C. (1987): *The Intentional Stance*. Cambridge, MA: MIT Press.

Downes, S. (2000): 'Truth, Selection and Scientific Inquiry', *Biology and Philosophy* 15, 425-442.

Garcia, J., et al. (1972): 'Biological Constraints on Conditioning', in Black and Prokasy (eds.), *Classical Conditioning: Current Research and Theory,* New York: Apppleton-Century Crofts.

Godfrey-Smith, P. (1996): *Complexity and the Function of Mind in Nature,* New York: Cambridge University Press.

Plantinga, A. (1993): *Warrant and Proper Function*, New York: Oxford University Press.

Plantinga, A. (2002): 'Naturalism Defeated', in Beilby (ed.) *Naturalism Defeated: Essays on Plantinga's Evolutionary Argument Against Naturalism*, New York: Cornell University Press.

Ramsey, W. (2002): 'Naturalism Defended', in Beilby (ed.) *Naturalism Defeated: Essays on Plantinga's Evolutionary Argument Against Naturalism*, New York: Cornell University Press.

Reed, T. M. (1997): 'Evolutionary Skepticism', *International Journal for Philosophy of Religion* 42, 79-96.

Stein, E. (1996): *Without Good Reason*, Oxford: Clarenden Press.

Sterelny, K. and Griffiths, P. (1999): *Sex and Death: An Introduction to Philosophy of Biology*, Chicago: University of Chicago Press.

Williams, G. C. (1966): *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*, Princeton, NJ: Princeton University Press.