

JAMES SAGE

TRUTH-RELIABILITY AND THE EVOLUTION OF HUMAN
COGNITIVE FACULTIES

I. INTRODUCTION

Many philosophers have claimed that evolutionary theory provides good reason to believe that human cognitive faculties are reliable. Specifically, they claim that reliable cognitive faculties enjoy a selective advantage over unreliable cognitive faculties. Thus, it is concluded that natural selection favors cognitive faculties that are capable of fulfilling the epistemic goal of having mostly true beliefs.

In this paper, I argue that the fact that human cognitive faculties have an evolutionary origin provides no reason to accept the claim that they are reliable with respect to generating true beliefs. While evolution might warrant the claim that human cognitive faculties are adaptive (or fitness-enhancing), the claim that human cognitive faculties reliably generate true beliefs is not supported by an appeal to evolution.

I begin by constructing an argument linking natural selection with reliability. I argue, however, that such an argument fails because it exploits an ambiguity between two distinct notions of reliability: fitness-reliability and truth-reliability. While it is plausible to think that our cognitive faculties are fitness-reliable, our epistemic goals require them to be truth-reliable. After discussing reasons for doubting whether true beliefs are more adaptive than false beliefs, I conclude that natural selection provides no reason to suppose that human cognitive faculties are truth-reliable.



Philosophical Studies 00: 95–106, 2003.

© 2003 Kluwer Academic Publishers. Printed in the Netherlands.

PDF-OUTPUT

WEB2C PDF-OP Disk, CP VICTORY: PIPS No.: 5151188 (philkap:humsfam) v.1.2
philpa25.tex; 2/10/2003; 18:14; p.1

II. EVOLUTIONARY RELIABILISM

The evolutionary reliabilist claims that evolution by natural selection shows that human cognitive faculties are reliable. Evolutionary reliabilism is the conjunction of four theses:

- K S knows that p only if S's true belief that p is generated by reliable cognitive faculties.
- R Human cognitive faculties are reliable.
- N Human cognitive faculties result from evolution by natural selection.
- Cond The conditional probability of R given N is greater than the probability of R.

In order to have reason to think that we meet the conditions for knowledge mentioned in K, we must have reason to accept R. In the face of this skeptical challenge, the evolutionary reliabilist tries to argue for R by appeal to scientific findings. If any scientific finding can support R, it is N. So, what needs to be argued is that accepting N provides good reason to accept R. Here, then, is a preliminary argument that connects natural selection with reliability:

1. Natural selection favors traits that increase an organism's inclusive fitness.
2. If an organism has cognitive faculties at all, it is more conducive to inclusive fitness to possess reliable cognitive faculties than to possess unreliable cognitive faculties.
3. Hence, natural selection favors reliable cognitive faculties over unreliable cognitive faculties.
4. If cognitive faculties evolved in some species during a long period of natural selection, then the cognitive faculties possessed by members of that species are reliable.
5. Human cognitive faculties result from a long period of evolution by natural selection [N].
6. Hence, human cognitive faculties are reliable [R].

I take it that premises 1 and 5 are basic claims of evolutionary biology. 3 follows from 1 and 2. 4 follows from 2 and 3. 6 follows from 3, 4 and 5. So, the controversial part of the argument is 2, which is the claim that it is more conducive to inclusive fitness to possess reliable cognitive faculties than to possess unreliable cognitive faculties. Is this claim reasonable?

Before we assess 2, we must understand how cognitive faculties might increase an organism's inclusive fitness. First, an organism's behaviors increase its fitness. A cognitive organism's behaviors are caused, in part, by the beliefs held by that organism, and the beliefs held by an organism are generated by the organism's cognitive faculties. These connections are a plausible way to link an organism's cognitive faculties with its fitness.

Premise 2 asserts that for cognitive organisms, it is more conducive to inclusive fitness to possess reliable cognitive faculties than to possess unreliable cognitive faculties. But 2 is true only if reliable cognitive faculties are more likely to increase an organism's inclusive fitness than are unreliable cognitive faculties.

Now, if by "reliable" all that is meant is that the cognitive faculty in question generates beliefs that reliably cause certain kinds of behaviors, and these behaviors (finding food, avoiding predators) increases an organism's inclusive fitness, then 2 is acceptable and the argument succeeds. But, if reliability means nothing more than reliably increases an organism's inclusive fitness, then lots of traits (both cognitive and non-cognitive) can be reliable with respect to fitness: the zebra's stripes are reliable, and so are the webbed feet of Mallard ducks. Moreover, a cautious cognitive faculty that "over detects" dangerous predators (frequently generating the false belief that a predator is nearby) may generate an abundance of false beliefs, though it may turn out to be adaptive because these false beliefs increase an organism's inclusive fitness.

Based on the preliminary argument, then, it is unwarranted to conclude that human cognitive faculties are reliable in the sense that they generate mostly true beliefs. And it is with this notion of reliability that epistemologists are primarily concerned. What is established is that human cognitive faculties reliably cause behaviors that increase inclusive fitness. This suggests that the argument trades on an ambiguity in the notion of reliability.

III. TWO SENSES OF RELIABILITY

The notion of reliability has two distinct senses. First, there is the biological sense in which a trait is reliable because it increases an organism's fitness. Call this "fitness-reliability." Such "fitness-

reliable” traits can be either cognitive or non-cognitive. There is another sense of reliability that epistemologists typically employ. Call this “truth-reliability.” A cognitive faculty is “truth-reliable” just in case it reliably generates true beliefs. So, *fitness-reliability* is a property of a trait that reliably contributes to fitness; *truth-reliability* is a property of a cognitive faculty that reliably generates true beliefs. We can now afford more precision regarding the claim that human cognitive faculties are reliable:

R_F Human cognitive faculties are fitness-reliable.

R_T Human cognitive faculties are truth-reliable.

Any epistemologist persuaded by the preliminary argument has failed to distinguish these two senses of reliability. The argument establishes R_F but fails to establish R_T , which is what the evolutionary reliabilist needs to be established. This distinction requires modification of K:

K^* S knows that p only if S’s true belief that p is generated by truth-reliable cognitive faculties

If we are to have reason to believe that we can meet our epistemic goals, we must have reason to believe R_T . The evolutionary reliabilist must now argue for R_T by appeal to N. How might this argument for R_T proceed?

The first option (a) keeps the preliminary argument above, but then argues for a connection between R_T and R_F . The second option (b) offers a new argument that utilizes truth-reliability throughout the argument.

The first option (a) argues for a connection between R_F and R_T via one of two strategies. The first strategy tries to deny the distinction between fitness-reliability and truth-reliability. This strategy is implausible, however, since numerous non-cognitive traits are fitness-reliable. Because non-cognitive traits do not generate outputs that carry truth values (i.e., beliefs), it follows that non-cognitive traits cannot be truth-reliable. Since non-cognitive traits *can* be fitness-reliable, the distinction between fitness-reliability and truth-reliability remains intact. So the first strategy under option (a) fails.

The second strategy under option (a) argues that the following conditional is true: if a cognitive faculty is fitness-reliable then it

is truth-reliable.¹ It is unreasonable for us to accept this conditional because we have good reason to believe that some fitness-reliable cognitive faculties are not truth-reliable: for example, highly cautious belief-forming strategies can generate adaptive yet false beliefs. Because fitness-reliable cognitive faculties can fail to be truth-reliable, the conditional fails. So the second strategy under option (a) fails. And thus option (a) fails altogether.

The second option (b) offers a new argument, similar to 1 thru 6 above. This new argument must establish R_T , but can appeal only to the connection between natural selection and truth-reliability. In what follows, I shall explore, and reject, this reconstructed argument.

IV. CONNECTING TRUTH-RELIABILITY WITH NATURAL SELECTION

Before turning to the reconstructed argument, let us first identify the general intuitive appeal behind attempts to provide reason for R_T based on N . The intuition is articulated (though not endorsed) by Feldman (1988, p. 218) in what he calls a “tempting argument”:

[I]f a being has beliefs at all, it is better (that is, more conducive to survival) for it to have true beliefs than false beliefs. True beliefs about where one’s food is are more helpful for finding food, and surviving, than are false beliefs. Similarly, true beliefs about where one’s predators are and how to escape them are more survival enhancing than false beliefs about these matters. So, natural selection is likely to select for believers that have mostly true beliefs. The best way, perhaps the only way, for believers to have mostly true beliefs is for them to have reliable belief-forming mechanisms or strategies. Reliable mechanisms or strategies are ones that lead mostly to true beliefs. Hence, natural selection will select believers that have reliable belief-forming mechanisms.²

Perhaps this “tempting argument” is what Quine had in mind when he claimed that “creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die out before reproducing their kind” (1969, p. 126). We find Daniel Dennett asserting that “Natural selection guarantees that *most* of an organism’s beliefs will be true, *most* of its strategies rational” (1987, p. 75). And before rejecting evolutionary theories of content, Jerry Fodor claimed that “Darwinian selection guarantees that organisms

either know the elements of logic or become posthumous” (1981, p. 121).

The intuition that these philosophers share is this: if organisms had largely *false* beliefs about their world, then they would fail to navigate their world successfully, and hence they would be unlikely to survive and reproduce. Competing organisms, who hold mostly true beliefs about their environment, will have a selective advantage over those who hold false beliefs. And over time, natural selection will eliminate the less fit, leaving just those organisms who manage to generate mostly true beliefs. Our human ancestors, it is thought, were among the survivors of this selection process. THEY were true believers. WE are their descendants.

So much for intuitions. Here is a reconstructed argument connecting natural selection (N) with truth-reliability (R_T):

- B1. Natural selection favors traits that increase an organism’s inclusive fitness.
- B2. If an organism has beliefs at all, it is more conducive to inclusive fitness to possess true beliefs than false beliefs.
- B3. Hence, natural selection favors true beliefs over false beliefs.
- B4. The best (perhaps only) way to generate mostly true beliefs is to possess truth-reliable cognitive faculties.
- B5. Hence, natural selection favors truth-reliable cognitive faculties over cognitive faculties that are not truth-reliable.
- B6. If cognitive faculties evolved in some species during a long period of natural selection, then the cognitive faculties possessed by members of that species are truth-reliable.
- B7. Human cognitive faculties result from a long period of evolution by natural selection [N].
- B8. Hence, human cognitive faculties are truth-reliable [R_T].

Premises B1 and B7 are basic claims of biological evolution. B3 follows from B1 and B2. B4 is plausible for present purposes.³ B5 follows from B3 and B4. B6 follows from B4 and B5. B8 follows from B5, B6, and B7. So the crucial step in the argument is B2.

Before assessing B2, it will be valuable to point out a general objection to this kind of argument: truth-reliable cognitive faculties

may never have been *available* in human evolutionary past. If truth-reliable cognitive faculties were never randomly generated (say, by mutation or genetic drift), then they could not be retained by natural selection.⁴ Just because cognitive faculties constructed from fiber optics would have been (and would be now) selectively advantageous, this cannot guarantee that cognitive faculties *are* constructed from fiber optics. Cognitive faculties made from fiber optics had to be available in order to be naturally selected. Similarly, truth-reliable cognitive faculties may never have been available in human history, so appealing to the supposed selective advantage of truth-reliable cognitive faculties cannot *guarantee* that human cognitive faculties are truth-reliable.

The availability objection is sufficient to undermine the claim that natural selection *guarantees* that cognitive faculties are truth-reliable. The availability objection is not fatal for the evolutionary reliabilist who must show that N provides good reason to accept R_T. A guarantee is not what is sought.⁵

What is wrong with the reconstructed argument? The point of contention, for the purposes of this paper, is premise B2. But B2 is true only if the following assumption is true:

- A True beliefs are more likely to increase an organism's inclusive fitness than are false beliefs.

Now, if we accept A, then we have good reason to accept B2, and the reconstructed argument succeeds (so long as we grant the other premises). The success of the reconstructed argument, therefore, depends on the plausibility of A. In the next section, I provide reasons to doubt A.

V. REASONS TO DOUBT A

What counts in favor of *accepting* A is the intuition articulated by Feldman (quoted at length above). This intuition invites us to imagine cases in which having true beliefs about food and water and safe hiding places will lead to adaptive behaviors. True beliefs, the intuition suggests, will lead to behaviors that are conducive to fitness and so favored by natural selection. All this intuition shows, however, is that there are some situations in which having true

beliefs is conducive to fitness. It does not establish that true beliefs are more likely to increase an organism's inclusive fitness than are false beliefs, which is what the intuition needs to do in order to support A.

Now, let's look at how might we provide reason to *doubt* A. A number of strategies come to mind. The first strategy identifies *particular* false beliefs that happen to be adaptive. Stich (1990, p. 58) provides an example along these lines: you survive a plane crash because you had a false belief about your departure time and therefore missed the plane. However, the fact that there are *particular* false beliefs that increase an organism's inclusive fitness does not seriously call A into question.

Another strategy identifies *clusters* of false beliefs that increase an organism's inclusive fitness. Some such clusters might be religious systems of belief. Because these various religious systems are incompatible, it follows that many of them are false. Therefore, individuals who accept these systems hold many false beliefs that, nevertheless, lead to behaviors that increase an individual's inclusive fitness. Some of these beliefs and behaviors might include believing the following: that hard work and clean living will glorify God, that having numerous children pleases Allah, that cooperating others is a form of praising Yahweh, that the use of birth control is prohibited by God, and so on and so forth. So, inclusive fitness can be increased by holding clusters of false religious beliefs. The abundance of adaptive false beliefs gives us reason to doubt that true beliefs are more likely to increase an organism's inclusive fitness than are false beliefs. And this is reason to doubt A.

A third strategy identifies fitness-reliable cognitive faculties or belief-forming processes that *systematically generate* false beliefs. The Garcia effect⁶ is an example of a belief-forming process that generates many false beliefs that increase an organism's inclusive fitness. For example, an organism may hide because it believes falsely that a predator is nearby. Evolutionarily, it pays to have cautious belief-forming processes that "over detect" dangerous predators, especially when false beliefs carry little cost. There are numerous examples of fitness-reliable processes that systematically generate false beliefs and, therefore, are not truth-reliable: believing that all spotted mushrooms are poisonous (because consumption of

spotted mushrooms was once followed by illness); believing that all strangers of the same sex are untrustworthy (because a few strangers of the same sex have been untrustworthy); and so on.⁷ So, cautious belief-forming processes (say, those based on weak inductive generalizations) can systematically generate false beliefs, but still be fitness-reliable. This strategy shows that in some cases having processes that typically generate false beliefs can increase inclusive fitness. And this is reason to doubt A.

Beliefs about the colors of objects may provide another example of adaptive false beliefs that are generated systematically. While it may be cautious to believe that all spotted mushrooms are poisonous, this may result in the exclusion of an important food source, and therefore decreased fitness. So, it might be beneficial to distinguish brown mushrooms with white spots from white mushrooms with brown spots (the former may be associated with sickness; the latter not). An organism unable to distinguish safe mushrooms from poisonous mushrooms may avoid all mushrooms, perhaps at a great cost. An organism with beliefs about the colors of mushrooms might be able to track additional consistencies, say, linking sickness with brown mushrooms with white spots. Beliefs about the colored objects, therefore, would allow an organism to select white mushrooms with brown spots as a food source, and this may increase inclusive fitness. Some philosophers and physicists, however, claim that physical objects are not colored at all. If they are correct, then all of our beliefs, to the effect that objects *are* colored, turn out to be false.⁸ Even if these beliefs turn out to be false, still, as I have argued, they may increase an organism's inclusive fitness. In some cases, therefore, a belief's ability to increase an organism's fitness is independent of its truth value. Again, this strategy suggests that fitness-reliable processes or strategies can systematically generate false beliefs. If adaptive false beliefs are generated systematically, then this provides some reason to think that false beliefs are in fact likely to increase inclusive fitness. And this is reason to doubt A.

While more precise "detection" of predators and poisonous foods might generate fewer false beliefs about predators and poisonous foods, the biological cost of such precision might outweigh the benefits of generating mostly true beliefs. More precise detection requires additional biological resources. Those resources could be

used elsewhere to increase inclusive fitness, and so natural selection might favor cognitive organisms with cautious, less precise belief-forming processes that are not truth-reliable.⁹ This observation suggests another way to call A into question.

A fourth strategy, then, identifies the biological cost of having truth-reliable cognitive faculties. Truth-reliable cognitive faculties come at a high price: (i) the brain requires oxygen, calories, and cooling, (ii) calculating detailed inferences (even with minimal data) requires considerable time and concentration, (iii) accessing information from past experience requires extensive storage capacity and retrieval pathways, (iv) identifying relevant information requires multi-level sorting subroutines, (v) ranking desires and goals requires extensive deliberation and reflection, and (vi) utilizing “detectors” (and other perceptual inputs) requires precision and acuity. Each of these factors carries a significant biological cost. Since biological resources utilized by truth-reliable cognitive faculties could be used in other ways to increase inclusive fitness (ways that would confer an immediate benefit to the organism), it follows that natural selection may favor fitness-reliable cognitive faculties that are *not* truth-reliable. And again, this is reason to doubt A.

VI. CONCLUSION

I have provided various reasons to doubt A. Because there is reason to doubt A, B2 remains unsupported. Without support for B2, we have no reason to accept the evolutionary reliabilist’s reconstructed argument linking natural selection, N, with the truth-reliability of human cognitive faculties, R_T. Therefore, the reconstructed argument fails to show how N provides good reason to accept R_T. And without good reason to accept R_T, we have no reason to think that humans meet the conditions for knowledge mentioned in K*. Thus we have no reason to accept evolutionary reliabilism. In particular, we have no reason to think that evolutionary considerations support the claim that human cognitive faculties are truth-reliable. And without such reason, the evolutionary reliabilist fails to meet the skeptical challenge that knowledge is possible, or that knowledge is a common achievement.¹⁰

NOTES

¹ My thanks to Tom Reed for pointing out that the evolutionary reliabilist may claim that this formulation of the conditional is too strong. Rather, what could be asserted is: if a cognitive faculty is fitness-reliable then it is *likely* to be truth-reliable. For reasons discussed below, regarding assumption A, I think this probabilistic rendering remains problematic.

² It is worth noting that Feldman does not, in the end, agree with this line of thought. See Clarke (1996) for similar assertions and statements of intuitions.

³ I think it is contentious, but I won't challenge it here. See Feldman (1988).

⁴ For similar points, see Sober (1981), Lycan (1988), and Stein (1996).

⁵ I think the availability objection can still create difficulties for the evolutionary reliabilist. I pursue such issues elsewhere.

⁶ See Garcia et al. (1972), Stein (1996, pp. 190–197), and Stich (1990, pp. 61–63).

⁷ These are examples of using weak induction to form beliefs. While weak induction might not be approved by logicians or Bayesians as a basis for forming beliefs (because, in part, such a method generates lots of false beliefs), it is not at all clear that the systematic generation of cautious false beliefs based on weak induction is necessarily maladaptive. Natural selection is not necessarily guided by the norms of Bayesian probability.

⁸ See Hall (1996).

⁹ In some cases, these processes may also be “fast and frugal.” See Gigerenzer et al. (1999) for a discussion of fast and frugal cognitive heuristics. While Gigerenzer uses “fast and frugal” to identify cognitive heuristics that are nevertheless truth-reliable (they generate truths despite taking shortcuts), I remain open to the view that such cognitive heuristics are not necessarily truth-reliable, though they are fitness-reliable.

¹⁰ I would like to thank Aaron Holland, Ram Neta, Lex Newman, and Tom Reed for comments on earlier drafts of this paper.

REFERENCES

- Clarke, M. (1996): ‘Natural Selection and Indexical Representation’, in M. Marion and R.S. Cohen (eds.), *Quebec Studies in the Philosophy of Science II*.
- Dennett, D. (1987): *The Intentional Stance*, Cambridge, MA: MIT Press.
- Feldman, R. (1988): ‘Rationality, Reliability and Natural Selection’, *Philosophy of Science* 55.
- Fodor, J. (1981): ‘Three Cheers for Propositional Attitudes’, in *Representations*, Cambridge, MA: MIT Press.
- Garcia, J. et al. (1972): ‘Biological Constraints on Conditioning’, in A. Black and W. Prokasy (eds.), *Classical Conditioning*, Hillsdale, NJ: L. Erlbaum Associates.

- Gigerenzer, G., Todd, P.M. and ABC Research Group (1999): *Simple Heuristics That Make Us Smart*, New York: Oxford University Press.
- Hall, R. (1996): 'The Evolution of Color Vision Without Colors', *Philosophy of Science* (Proceedings) 63, S125–S133.
- Lycan, W. (1988): *Judgment and Justification*, Cambridge: Cambridge University Press.
- Quine, W.V.O. (1969): *Ontological Relativity and Other Essays*, New York: Columbia University Press.
- Sober, E. (1981): 'The Evolution of Rationality', *Synthese* 46, 95–120.
- Stein, E. (1996): *Without Good Reason*, Oxford University Press.
- Stich, S. (1990): *The Fragmentation of Reason*, Cambridge, MA: MIT Press.

Department of Philosophy
University of Wisconsin-Stevens Point
Stevens Point, WI 54481
USA
E-mail: jsage55@attbi.com