A Flaw in the Stich-Plantinga Challenge to Evolutionary Reliabilism

Michael J. Deem
Duquesne University

1 Introduction

Did selective pressures shape in humans over the course of their evolutionary history cognitive faculties that reliably produce true beliefs? It seems plausible that true beliefs about, say, the location of food sources, the relative proximity of predators, or the presence and availability of potential mates would have provided an adaptive advantage to individuals whose cognitive machinery reliably produced them, enabling these individuals to out-reproduce conspecifics whose beliefs about these environmental features were mostly false. Accordingly, it might be thought, natural selection would have culled from the population the latter while maintaining and amplifying in the former those cognitive faculties responsible for producing true beliefs.

The view that selective pressures would likely have favored cognitive reliability in humans is commonly called 'Evolutionary Reliabilism' (ER). Several philosophers endorse ER, though there are differences among them regarding precisely which cognitive faculties were targeted by natural selection and why.[1] Nonetheless, the common argumentative thread is that natural selection would likely have favored reliable cognitive processes – that is, processes that

---

[1] For example, some philosophers claim that natural selection would have targeted particular cognitive capacities, such as those that produce accurate inductive beliefs and predictions about the environment (e.g., Quine 1969). Others argue that belief content serves like a map or guidance system, and accurate beliefs would have directed behaviour in adaptive ways (e.g., Ramsey 2002; Boutler 2007). Others suggest that there was a less direct effect by selective pressures on belief content, but given certain conceptual constraints on the link between content and behaviour, selection would have favored content that tended to correspond to adaptive behaviour (e.g., Law 2012; perhaps Ruse 2012).

tend to produce a high proportion of true beliefs – over unreliable cognitive processes, because the former would have been more conducive to survival and reproductive success than the latter.

While ER enjoys some plausibility, the view has been subject to serious criticism. Perhaps the most familiar critiques of ER are those tendered by Stephen Stich and Alvin Plantinga. While Stich and Plantinga embark on divergent argumentative paths toward different philosophical goals -- Stich seeking to vindicate a particular brand of epistemic pragmatism, Plantinga aiming to undermine naturalism -- each develops along the way a sub-argument for a strikingly similar claim. Both argue that there is no good reason to think that natural selection would favor reliable belief-producing faculties in humans. Stich and Plantinga not only arrive at this same conclusion, but each relies on the same crucial premiss; namely, that natural selection is indifferent to true belief.[2]

In this article, I argue that this premiss is both imprecise and too weak to support Stich's and Plantinga's desired conclusion. I organize my discussion in the following way. In §2 and §3, I describe and clarify Stich's and Plantinga's main arguments against ER. In §4, I identify and clarify the key premiss that their arguments share, and I show that it cannot generate their arguments' conclusion regarding the relation between selective forces and human cognitive faculties. Their challenge to ER, I conclude, fails. Finally, in §5, I suggest what is required of a successful defence of ER.

---

[2] In their critiques of ER, Stich and Plantinga take the relevant notion of belief to be that of a psychological state with propositional content, and a belief is true or false in virtue of that content (Stich 1990: 22-23, 61-62; Plantinga 2011b: 322). If a belief that p is true, it is true because the proposition p is true. Neither expressly links the truth-conditions of beliefs to adaptive behaviour, as some philosophers have done in developing naturalized and evolutionary epistemologies (Godfrey-Smith 1996 contains a very helpful discussion of such success-linked semantic theories).

2 Stich's Challenge to ER

I begin with Stich's discussion of the evolution of human cognitive faculties (Stich 1990). Stich begins with a relatively uncontroversial assumption about biological fitness and inferential systems; namely, all other things equal, one inferential system will be favored by natural selection over another system if, on balance, the former is more conducive to survival and reproductive success than the latter.

Stich then sketches the following evolutionary scenario in which a reliable inferential system might be selected *against* and in which a less reliable one might be favored. Suppose an organism lives in an environment with high gastronomical variability, and consider two possible inferential strategies for determining which foods are poisonous in this environment. One strategy is quick and risk-averse in making inferences about which foods are poisonous, and the other is more calculative and deliberative. While the risk-aversive system will likely produce a fair number of false positives about, say, which berries and tubers are toxic, it is unlikely to yield a false negative, which could be fatal for the individual. As long as the risk-aversive system does not compromise the individual's ability to receive adequate nutritional support or to reproduce, several false positives may well be less costly than the single false negative yielded by a more cognitive complex, deliberative system.

Moreover, the risk-aversive system may incur a much smaller biological maintenance cost in terms of nutrition, execution time, and energy than would a more cognitively complex and demanding system, allowing the organism to devote these resources toward, say, foraging or reproductive pursuits. The higher biological maintenance cost of the complex, reliable system, then, might be a comparatively less efficient evolutionary solution to the gastronomical problems posed by this ancestral environment.

The risk-averse strategy, then, might on balance confer greater fitness benefits on individuals than the latter within this hypothetical environment. Thus, Stich:

> [T]he unreliable, error-prone, risk-aversive strategy may well be favored by natural selection. For natural selection does not care about truth; it cares only about reproductive success. And from the point of view of reproductive success, it is often better to be safe (and wrong) than sorry. (1990: 62)

Similar hypothetical evolutionary scenarios might also show that a more risk-aversive, but less reliable inferential system would provide greater adaptive payoffs than its reliable counterpart.[3] It does not follow, then, from the assumption that selective pressures targeted human cognitive faculties that selection shaped such faculties to be reliable. Stich concludes that there is little basis for thinking that ER is correct.

3 Plantinga's Challenge to ER

I turn now to Plantinga's evolutionary argument against naturalism (EAAN), and specifically to its first premiss.[4] This premiss states that the probability (P) that our cognitive faculties reliably produce true beliefs (R), conditioned on the truth of naturalism (N) and evolutionary theory (E), is low or inscrutable -- or P(R/N&E). '(N)' stands for what Plantinga

---

[3] Sage (2004) suggests that unreliable cognitive processes might be more adaptive than reliable processes across a broad set of ancestral circumstances in which cognitive traits would have mattered to survival and reproductive success. False beliefs about cultural practices (e.g., religious doctrines) or processes that systematically produce false beliefs (e.g., colour-attribution to objects in one's environment) might have been adaptive in particular ecological and social contexts.

[4] The EAAN has undergone several modifications over the past two decades, the most recent permutation being found in Plantinga 2011a and Plantinga 2011b. I will focus on this latest version, since Plantinga regards it as 'the official and final version' (2011b: 310n4).

calls 'metaphysical naturalism'. Though he does not offer a precise characterization of (N), he takes (N) to imply that 'there is no such person as God or anything at all like God -- or if there is, this being plays no causal role in the world's transactions' (2011a: 435). Now, one might object that it is unfair to Plantinga to group his argument for this premiss along with Stich's argument against ER, since Plantinga clearly states that that it is the *conjunction* of (E) and (N), and not just simply (E), that renders the probability of cognitive reliability low or inscrutable. However, it seems that Plantinga is committed to the view that P(R/E) is low or inscrutable. I suggest this for two reasons. First, Plantinga contends that there is little reason to think that the '*unguided* evolution' of human cognitive capacities (an implication of (N&E)) would result in a tight link between belief and truth, while 'evolution just as such' (i.e., (E)) does not inspire such doubt (2011b: 308-09). However, as Plantinga himself notes, contemporary evolutionary theory generally adheres to *methodological naturalism*, which he characterizes as the view that 'in science, we should proceed as if the supernatural is not given' (2011b: 170). But assuming evolutionary theory proceeds *as if* the process of evolution by natural selection were an unguided process of genetic changes and selection against maladaptive traits, it is not clear that the adoption of (N) would in any way modify the content of (E). Evolutionary processes would look the same on (E) as on (N&E).

Second, Plantinga appears to hold that it is only when (E) is conjoined to a particular brand of *theism* that we have good reason to think that human cognitive faculties are reliable (2011b: 339n29). This, then, suggests that (E) alone does not provide good reason to think that human cognitive faculties evolved to be reliable, quite apart from what (N) entails about supernatural intervention in evolutionary processes. Theism might add guidance to (E), but (N)

subtracts nothing from (E)'s content. Thus, it seems that Plantinga is committed to the view that P(R/E) is low or inscrutable.

Let us now consider the rationale behind the first premiss of the EAAN. To support the premiss, Plantinga employs a strategy similar to that which Stich utilizes in his argument against ER. Plantinga offers two hypothetical evolutionary scenarios that purportedly show that natural selection does not necessarily target the content of beliefs. In the first scenario, he asks us to consider when belief content might have emerged on the evolutionary timescale. Assuming that a belief consists of a structure of neurophysiological (NP) properties and representational content that is consequential on that structure, do we have reason to think that selection would target the content? At some point in the evolutionary process, NP properties in some organisms passed a certain threshold of organization and complexity, giving rise to rudimentary belief-like states. Presumably, at this evolutionary stage there already was a tight causal link shaped by selection between these organisms' NP properties and their behaviour such that certain NP configurations triggered by environmental features reliably generated behaviour that was conducive to survival and reproductive success. What, Plantinga asks, is the likelihood that this rudimentary representative content would be true, or that this content would be relevant to the fitness of such organisms? His response:

> All that's required for survival and fitness is that the neurology cause adaptive behavior; this neurology also determines belief content, but whether or not that content is true makes no difference to fitness. (2011b: 327)

At most, this scenario shows that there might not have been an *initial* link between accurate or true rudimentary representative content and adaptive behaviour. But it does not provide any reason to think that selection would not favor true belief at some *later* stage of the evolutionary process, as organisms' detection systems and cognitive processing became more complex.[5] It turns out, then, that Plantinga's first evolutionary scenario presents little challenge to ER.

Plantinga's second hypothetical evolutionary scenario, however, is more relevant to the question of whether selection favored true belief among more cognitively complex organisms. Imagine a race of organisms bearing cognitive traits much like our own (e.g., capacities for belief-formation, inference-making, etc.). Suppose selective forces shaped these traits. Any given belief of one of these cognitively-complex organisms will consist of a structure of NP properties and representational content. How would these belief states link to features of the environment -- say, to the presence of a predator? Assuming that the NP structure of the belief is of the sort that arises reliably in the actual presence of a predator, do we have any reason to think the content of the belief would be true, that is, would correspond to that which triggers the NP structure? Plantinga:

> The fact that these creatures have survived and evolved, that their cognitive equipment
>
> was good enough to enable their ancestors to survive and reproduce -- that fact would tell

---

[5] ER takes reliable belief-producing faculties to be an adaptation, evolutionarily shaped on account of their capacity to produce a preponderance of true beliefs. ER, then, is committed to the view that belief content plays or once played a causal role in organismic behaviour; otherwise, belief content would not have been targeted by selective pressures. The connection, then, between belief content and behaviour is not incidental on ER, as if the relation between content and behaviour were like that of the sound of a heart beating and a heart's function of pumping blood. What prevents Plantinga's argument from begging the question against ER is that he does not assert that belief content *never* came under selective pressure, but only asserts that the *emergence* of belief content was incidental to the evolutionarily fixed link between NP structures and behaviour. Still, even this assertion is controversial, and he provides no argument for the view that the evolutionary emergence of representational content within organisms' detection systems was purely incidental.

us nothing at all about the truth of their beliefs or the reliability of their cognitive

faculties. (2011b: 331)

Plantinga thus infers from this scenario that, given (E), the beliefs of this hominin-like organism

are just as likely to be true as they are to be false.

Like Stich, Plantinga takes his hypothetical evolutionary scenarios to show that there is

no necessary connection between selective pressures and cognitive reliability. Further, if we can

imagine scenarios in which an individual's biological fitness either is not compromised by the

production of false beliefs or is enhanced by the production of false beliefs, then natural selection

is indifferent to true belief. And if selection is indifferent to true belief, then ER is undermined.

Is the Stich-Plantinga move from abstract, hypothetical evolutionary cases to conclusions

about whether we have reason to think reliable human cognitive faculties evolved warranted? I

argue in the following section that it is not.

4 The Flaw

From the foregoing considerations, it is clear that Stich's and Plantinga's arguments

against ER rely on similar conceptions of the relationship between natural selection and the

cognitive faculties of organisms. They express this conception in virtually equivalent ways:

Stich: '[N]atural selection does not care about truth; it cares only about reproductive

success.' (1990: 62)

Plantinga: '[N]atural selection is interested, not in truth, but in appropriate behavior.'

(2011b: 316)

Call the claim that natural selection is not interested in truth the INDIFFERENCE CLAIM.

While Stich and Plantinga succeed in showing that natural selection, considered in isolation from any particular population and any specific environment, does not *necessarily* favor cognitive reliability, neither considers the more relevant question to debates over ER of whether the *historical* evolution of humans in their ancestral environments provides any reason to think cognitive reliability likely was favored by selective pressures. In the absence of such considerations of the evolutionary history and ecological conditions in which human cognition evolved, how can the INDIFFERENCE CLAIM be justified, and what bearing might it have on the question of whether selection historically favored cognitive reliability in humans?

It turns out that the INDIFFERENCE CLAIM, which serves as perhaps the most crucial premiss in the Stich-Plantinga challenge to ER, can do very little work. This is because, strictly speaking, selective forces considered in isolation from historical populations are as indifferent to truth as they are to showy plumage, pointed beaks, and self-sacrificial behaviour on behalf of other members of the group. There is no necessary connection between natural selection and these traits. Does this then show that selection actually was indifferent to them? It seems not; with respect to historical populations (as opposed to hypothetical populations) we do have good reason to think that natural selection favored these latter three traits in peacocks, cactus finches, and ants, respectively.[6] We suspect that, over the course of these species' evolutionary histories,

---

[6] One might object that we know that organisms actually have these traits, while the presence of cognitive reliability in humans is precisely what is at stake in this debate. Comparing cognitive reliability to traits like pointed beaks or showy tails, the objection might go, is question begging. But this objection fails. If we are convinced that human beings do not have reliable belief-forming faculties in the first place, then any evolutionary argument for or against

having these traits likely facilitated adaptive behaviour for individuals of these populations within their particular ancestral environments. To see the flaw in the Stich-Plantinga challenge to ER, notice what happens when we substitute any one of these adaptations into the INDIFFERENCE CLAIM:

> Natural selection does not care about *showy plumage*, but only about adaptive behavior.

Here, we get the same result for showy plumage as Stich and Plantinga do for cognitive reliability. But this certainly is not the result we want for explaining why peacocks tend to have showy plumage. It appears something has gone wrong in the Stich-Plantinga challenge to ER.

The problem is that the challenge relies on a *species-neutral* claim about what natural selection favors in order to derive a *species-specific* conclusion about the evolution of a particular trait. From this species-neutral viewpoint, the most we can say is that selection favors traits that confer adaptive advantage to individuals. But from this viewpoint alone, we cannot determine which *specific* traits might have been the ones that facilitated adaptive behaviour over the evolutionary history of an actual population, and were therefore likely favored by selection.

Even if we construct hypothetical evolutionary scenarios in which some given trait would likely not be favored by selection -- as Stich and Plantinga do in order to test the adaptedness of cognitive reliability -- we still could not determine whether or not that trait *was* favored. Consider once more our peacocks. We can easily imagine evolutionary scenarios in which, say, drab tails on peacocks would have been more adaptive than their showy counterparts -- say, in

---

cognitive reliability in humans is moot. There is no point in providing an evolutionary explanation for a trait that does not or did not exist in a population. In order to provide the sort of evolutionary perspective that the Stich-Plantinga challenge demands, cognitive reliability must be posited as a trait that is or was once present in the hominin lineage.

some hypothetical environment in which peacock predators are far more attracted to gaudy plumage than are peahens. In this scenario, gaudiness clearly does not pay. Does the fact that we can imagine such an evolutionary scenario in which showy plumage would be selected against show us that natural selection was indifferent to showy plumage over the course of the actual evolution of peacocks? I don't see how. In the same way, simply imagining cases in which reliable cognitive processes would not be as favored as, say, risk-aversive, fast-and-frugal inferential systems, or simply imagining hominin-like populations in nondescript environments, will not tell us whether selection actually favored cognitive reliability in the historical evolutionary lineages of hominins.[7]

The INDIFFERENCE CLAIM, therefore, contributes very little to our understanding of the evolutionary emergence and maintenance of cognitive reliability in humans. As the peacock example illustrates, thinking about selective pressures while abstracting from historical populations and their ancestral environments will not resolve the question of whether selection actually or likely favored cognitive reliability in humans any more than it will for showy plumage in peacocks. Thus, the Stich-Plantinga challenge, because it is primarily driven by the INDIFFERENCE CLAIM, fails to undermine ER.

5 Concluding Remarks

I have argued that the Stich-Plantinga challenge to ER fails. The key premiss of the challenge, the INDIFFERENCE CLAIM, is not justified by Stich's and Plantinga's appeals to abstract, hypothetical evolutionary scenarios in which cognitive reliability does not play an

---

[7] There is broad agreement that the peacock's showy plumage is the result of sexual selection pressures, due to heightened competition among peacocks for mates and selectivity in mate choice among peahens. While some researchers suggest that sexual selection played a significant role in shaping human cognitive capacities (e.g., Miller 2000), my use of the peacock example should not be understood as an endorsement of this view. Sexual selection is one mode of natural selection, and my argument here is neutral with respect to which mode(s) of selection (e.g., sexual, group) were primarily responsible for shaping human cognitive capacities. Nothing turns on which specific adaptation is substituted for 'truth' in the INDIFFERENCE CLAIM.

adaptive role. At most, this challenge establishes the claim that natural selection does not *necessarily* favor cognitive reliability in humans. But this modal claim, while perhaps philosophical interesting, tells us nothing with respect to whether selection historically favored cognitive reliability in humans.

Now, Stich and Plantinga might reply that their aim is only to show that cognitive reliability in humans is not the necessary outcome of selection pressures on human cognition. But, since ER does not entail a necessary connection between selective pressures and cognitive reliability, the proponent of ER can endorse their claim. Again, we cannot validly infer, strictly on the basis of this modal claim, a conclusion about whether or not there is reason to think that cognitive reliability historically played an adaptive role over the course of human evolution.

If my argument here is successful, it wards off a prominent challenge to ER. But my negative argument is not itself sufficient to justify ER. What then is needed to provide a successful defence for ER? As this paper suggests, we need to rely on more than conceptual considerations and purely hypothetical evolutionary scenarios. In order for a defence of ER to be successful, it needs also to include consideration of the environmental and social features that likely were present in humans' early ancestral environments, how some of these features would have generated selective pressures on human cognition in the direction of reliability, and whether there are phylogenetic constraints and homologies that shed light on the the evolutionary function of cognitive reliability. This requires philosophers to engage theoretical and empirical work in the biological sciences on the evolution and development of human cognition, and important philosophical work has already been done in this regard (e.g., Godfrey-Smith 1996; Sterelny 2003, 2012). The real challenge, then, for the proponent of ER is to provide a plausible,

biologically-informed account of why selective forces historically favored cognitive reliability

over the course of human evolutionary history.[8]

References

Boulter, S. J. 2007. The 'evolutionary argument' and the metaphilosophy of commonsense. *Biology and Philosophy* 22: 369–82.

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

Law, S. 2012. Naturalism, evolution and true belief. *Analysis* 72: 41–48.

Miller, G. 2000. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York: Doubleday.

Plantinga, A. 2011a. Content and natural selection. *Philosophy and Phenomenological Research* 83: 435–458.

Plantinga, A. 2011b. *Where the Conflict Really Lies: Science, Religion, and Naturalism*. Oxford: Oxford University Press.

Quine, W. V. 1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Ramsey, W. 2002. Naturalism defended. In *Naturalism Defeated?: Essays on Plantinga's Evolutionary Argument against Naturalism*, ed. J. Beilby, 15–29. Ithaca, NY: Cornell University Press.

Ruse, M. 2012. *The Philosophy of Human Evolution*. Cambridge: Cambridge University Press.

Sage, J. 2004. Truth-reliability and the evolution of human cognitive faculties. *Philosophical Studies* 117: 95–106.

Sterelny, K. 2003. *Thought in a Hostile World: The Evolution of Human Cognition*. Oxford: Blackwell.

Sterelny, K. 2012. *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, Mass: MIT Press.

Stich, S. P. 1990. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, Mass: MIT Press.