

2011

Plantinga, Kant, and Cognitive Reliability

Peter Zuk
Pepperdine University

Follow this and additional works at: <https://digitalcommons.pepperdine.edu/globaltides>



Part of the [Philosophy Commons](#)

Recommended Citation

Zuk, Peter (2011) "Plantinga, Kant, and Cognitive Reliability," *Global Tides*: Vol. 5, Article 2.
Available at: <https://digitalcommons.pepperdine.edu/globaltides/vol5/iss1/2>

This Humanities is brought to you for free and open access by the Seaver College at Pepperdine Digital Commons. It has been accepted for inclusion in Global Tides by an authorized editor of Pepperdine Digital Commons. For more information, please contact bailey.berry@pepperdine.edu.

“Plantinga, Kant, and Cognitive Reliability”

By **Peter Zuk**

ABSTRACT

This article applies Alvin Plantinga’s principle that self-defeat is inherent in any theory which gives the theorist reason to doubt her own cognitive faculties to Kant’s theory of perception, transcendental idealism. Because transcendental idealism excludes the possibility of knowledge about things in themselves, including the transcendental idealist’s own cognitive faculties, the theory is self-defeating under Plantinga’s argument. The article continues by arguing that Kant’s appeal to God as a way to vindicate the transcendental idealist’s cognitive faculties is problematic in several ways, and that the theory therefore cannot avoid self-defeat.

Alvin Plantinga has recently done substantial work on an Evolutionary Argument Against Naturalism (EAAN). It purports to show that belief in naturalism is self-defeating when held in conjunction with a belief in evolutionary theory, because upon reflection, the pair constitutes a defeater for belief in the reliability of the naturalist’s cognitive faculties. This in turn constitutes a defeater for all of the naturalist’s other beliefs, including belief in naturalism. The basic principle behind this argument – that self-defeat is inherent in any theory which gives the theorist reason to doubt her own cognitive faculties – can be employed as a powerful objection to Immanuel Kant’s theory of perception, transcendental idealism. It will be shown that transcendental idealism falls to this objection due to its exclusion of the possibility of knowledge about things in themselves, which includes the transcendental idealist’s own cognitive faculties as they are in themselves. It will further be argued that Kant’s appeal to God as a way to vindicate the transcendental idealist’s cognitive faculties is problematic in several ways, and that the theory therefore cannot avoid self-defeat. Because of this, Kant fails to achieve his goal of demonstrating a middle ground¹ with regard to the scope of human knowledge.

Plantinga’s reasoning goes something like this: we have a defeater for any belief or set of beliefs that entails a low or inscrutable probability of our having reliable cognitive faculties. This is so because the belief or set of beliefs was formed by the very thing it raises doubts about the reliability of. Let R stand for our having reliable cognitive faculties and P stand for (what we

take to be) the probability of this being the case. Thus, let “P of R” stand for the probability of R being true. Plantinga justifies the inference from a low P(R) to defeat of belief in R using the example of a man who has just ingested some psychoactive drug that renders the cognitive faculties of its victims unreliable in the vast majority of cases and does so in such a way that the change is undetectable to the victim. In such a situation, Plantinga says, the man clearly has a defeater for his belief in R, and thus, all of his other beliefs as well. They are all suspect because of the low chance of the faculties that produce and govern them working properly. And none of these beliefs can come to his aid because they all have their ultimate basis in his cognitive faculties, the reliability of which are precisely what is under scrutiny.² Similarly, we also have a defeater for belief in R in cases in which P(R) is inscrutable, that is, unknowable. Plantinga puts himself in the shoes of a Humean skeptic who is agnostic with regard to the origin of his cognitive faculties:

I have no idea what the probability of my faculties being reliable is, given the relevant facts about their origin and purpose. But then I have a defeater for my original belief or assumption that my faculties are in fact reliable. If I have or can get no further information about their reliability, the reasonable course for me is agnosticism with respect to R, giving it up, failing to believe it.³

It is clear that in both cases, the subject has a defeater for belief in R. The evolutionary process is likely to have given rise to reliable cognitive faculties only if it involved a filtering of false beliefs.⁴ Plantinga argues that the probability of *naturalistic* evolution including such a filter is either low or inscrutable, which entails a low or inscrutable value for P(R) itself. He supports this claim by pointing out that a belief can be adaptive without being true. The theory of evolution asserts that adaptive traits will tend to be passed on to future generations, and this would include cognitive faculties that form beliefs that aid in survival (regardless of whether or not those beliefs are true). The only naturalistic account of how our cognitive faculties came to be is evolution, and (Plantinga asserts) there is no good reason to suppose that unguided evolution would produce these faculties such that they were reliable. So the tension is not between naturalism and evolution per se, but rather between naturalism and its own general inability to provide good reason for thinking that our cognitive faculties *are* reliable. On this crucial point, it

is argued that naturalism fails to deliver. C.S. Lewis, whom Plantinga credits as the first thinker to point out naturalism's supposedly self-defeating character, explains:

A theory which explained everything else in the whole universe but which made it impossible to believe that our thinking was valid, would be utterly out of court. For that theory would itself have been reached by thinking, and if thinking is not valid that theory would, of course, be itself demolished. It would have destroyed its own credentials.⁵

The contention of Plantinga and Lewis is that naturalism is a theory of precisely this sort: one that excludes the possibility of validating our cognitive faculties. This is not to say that a good theory must prove itself (any such effort will always be circular), but rather that it must be self-consistent. Plantinga and Lewis' contention is that no theory which implies a low or inscrutable probability of our having reliable cognitive faculties can be consistent in this way.

Irrespective of how the objection fares against naturalism, its underlying principle is a valuable philosophical contribution. A theory that precludes an affirmation of the cognitive reliability of the theorist, either by providing evidence against it or by taking the question off of the table entirely, is an untenable one. To do the former is to assert that $P(R)$ is low and to do the latter is to assert that $P(R)$ is inscrutable, both of which defeat belief in R and thus the rest of one's beliefs. It is important to note that a theory is not self-defeating in this way simply because it fails to arrive at an affirmation of cognitive reliability, however. The theory of special relativity, for example, does not lend any support to the idea that human mental processes can get at truth. Yet of course one can still believe in the truth of special relativity without contradiction. The reason is that while the theory gives us no evidence in favor of cognitive reliability, neither does it give us any evidence against reliability, nor imply any kind of limitation on our part with regard to establishing it. Here $P(R)$ is inscrutable in a different sense, that is, solely *in the context of the theory*. The theory of special relativity simply has nothing to say about $P(R)$, so no epistemic limit is placed on our ability to ascertain it in other ways. It is not unknowable in principle, and that is the crucial distinction.

These considerations provide a formidable objection to Immanuel Kant's theory of perception, transcendental idealism. Arthur Schopenhauer succinctly characterizes the system as the

“distinction between the phenomenon and the thing in itself, and a recognition that only the phenomenon is accessible to us.”⁶ From this it follows that “we do not know either ourselves or things as they are in themselves, but merely as they appear.”⁷ According to Kant, we have access to subjective phenomena, but not the objective things in themselves that lie behind them. This is because we impose certain concepts on the world in order to understand it. Things such as space and time are not real features of objects, but features of our mental *representations* of objects. Or, at the very least, we could never know that they are real features of objects even if they are, because we can only form representations of what is given to us in experience. We can never know the world as it really is, Kant claims, because all of our experience of it is necessarily colored by the concepts we impose upon it. It is impossible to divorce our experiences from the lens through which we view them. Thus under transcendental idealism, we lack insight as to the nature of things in themselves. Kant says:

When I speak of objects in time and in space, it is not of things in themselves, of which I know nothing, but of things in appearance...Objects of the senses therefore exist only in experience.⁸

In light of our prior examination, however, such an assertion is problematic. This is because it limits our knowledge not merely to objects of experience, but to our mere *mental representations* of those objects. The objects themselves are unknown to us.⁹ As Kant himself admits, this includes the self. He says:

I, as an intelligence and *thinking* subject, cognize myself as an object [in] *thought*... like other phenomena, not as I am in myself, and as considered by the understanding, but merely as I appear.¹⁰

We can of course have self-knowledge in the same way we can have knowledge about objects. But such knowledge refers only to the representation given to us in experience (by mental states); it does not pertain to the true self. And if we cannot have knowledge about the real composition of the self, we cannot have knowledge about the reliability of our cognitive faculties. We can ascribe no value to P(R). It is inscrutable by definition given the truth of transcendental idealism. Thus Plantinga’s argument can be applied effectively to Kant’s system. Inherent in the very notion of transcendental idealism is the impossibility of discovering P(R), because we cannot have knowledge of our cognitive faculties or the processes that formed them as they are in themselves. While naturalists can deny the force of Plantinga’s argument against

them by arguing that the evolutionary process *would* include a filter of false beliefs, the transcendental idealist can make no such move because she can have no significant beliefs about things in themselves to appeal to in the first place. The most potent objections to Plantinga's original argument are thus unavailable to the transcendental idealist.

Insofar as an understanding imposes features on the world rather than perceiving ones that it really possesses, it is at best a sleeper conducting a rigorous scientific investigation on a dreamworld. It might come to know every 'fact' and 'principle' about the dream and yet remain ignorant as to the character of the real world, including its own nature. It is by this analogy that Kant's predicament becomes clear. It could never be established by such a process that the process itself is conducive to getting at truths about the real world, because there is no way to know whether these findings reflect reality or are themselves part of the dream. And if that is so then there is no basis on which to affirm the original assertion, namely, that we must necessarily view the world in a way analogous to this dream. Kant attempts to counter by claiming that what is ultimately objective and real is that which all of our experience has in common:

Objective validity and necessary universal validity (for everybody) are equivalent concepts, and though we do not know the object in itself, yet when we consider a judgment as universally valid, and hence necessary, we understand it thereby to have objective validity.¹¹

No sense of the word 'objective' is applicable here, however. Any reference to what the term denotes – facts about things in themselves independent of what we think – is excluded by transcendental idealism. It cannot mean mere agreement because no amount of subjective points of view could ever constitute an objective truth. We might arrive at universal agreement on all sorts of propositions and yet be no closer to knowing the truth, the actual fact of the matter. It must, then, refer only to claims made about our shared dream, and this profoundly limits our ability to make meaningful statements at all.

Kant himself saw the implications of transcendental idealism's exclusion of self-knowledge: a collapse into total skepticism. Rather than rejecting it, however, he sought refuge in pragmatic considerations. By positing certain (existent but unobservable) things in themselves, reason could be validated:

We must therefore think [of] an immaterial being, a world of understanding, and a Supreme Being... because in them only, as things in themselves, reason finds that completion and satisfaction, which it can never hope for in the derivation of appearance.¹²

But this is deeply problematic. A belief's convenience has no bearing on its truth. The fact that the existence of God, the soul, and a world conducive to inquiry would save transcendental idealism from collapse in no way entails their actual existence. Kant is correct that these things would allow for the "greatest possible use of reason"¹³ within the framework of transcendental idealism, but nothing about their practical value makes us any more justified in believing in them. Any use of reason we base on these kinds of propositions is permissible only if the propositions are true, and on Kant's view we have no way of obtaining genuine knowledge about that. So P(R) would still be inscrutable for the transcendental idealist, and she would still have a defeater for R and the rest of her beliefs with it, including her belief in transcendental idealism itself.

Furthermore, anyone who grants that beliefs can be held on pragmatic grounds need not subscribe to transcendental idealism a moment longer. Once it is allowed that beliefs be held on the basis of usefulness rather than evidential support, it is difficult to see why a system that denies the possibility of meaningful knowledge of the self and the external world would remain part of anyone's worldview. Kant shores up his theory in a non-epistemic way to prevent the unpalatable effects of skepticism. If this is allowable, there is no good reason why someone with a realist theory of perception cannot do the same to avoid the unpalatable effects of Kant's system.

Even if we suppose that Kant and Kant alone has the privilege of making this kind of move, serious questions still remain. Under transcendental idealism we cannot attribute properties to God as he really is, only as he is according to the apprehension we have come up with.¹⁴ This is problematic on a number of levels, the most obvious of which being that God is not given to us in direct experience. Without a robust experience of God it is unclear how we can have even an apprehension of him. Regardless, this is still not a tenable solution to the problem. Even allowing for an apprehension of God does not get us any closer to establishing the possibility of knowing

P(R). The transcendental idealist must still, in the words of Kant, “acknowledge that the Supreme Being is quite *inscrutable* and even unthinkable in any determinate way as to what it is in itself.”¹⁵ It follows that the positing of a mere representation of ‘God’ would not provide us with knowledge about our cognitive faculties as they really are. Rather, we would be required to posit God as he is in himself, which is impossible by definition under transcendental idealism. Our ability to do such a thing would *rule out* transcendental idealism.

Finally, suppose we grant all of Kant’s claims regarding God, including entailment of the all-important high P(R) such that belief in R is justified. He now seems very much committed to the claim that there exists a God concerned with providing us with conditions conducive to our obtaining knowledge. But transcendental idealism also commits Kant to a claim about the reliability of our *faculties of perception*. Let F represent the claim that our faculties of perception are reliable at apprehending things in themselves. Transcendental idealism hinges on the assertion that P(F) is low or inscrutable. Yet a low or inscrutable P(F) and the existence of this God are in tension: it seems quite unbecoming of a God concerned with helping us access knowledge to limit our faculties of perception so profoundly as the transcendental idealist is now committed to saying he has. Given the existence of a creator deity concerned with providing us with reliable cognitive faculties, it seems that we could expect to have reliable faculties of perception as well. It is reasonable to suppose that a God who is not a deceiver in one area would behave similarly in others (especially ones so closely related as perceptual faculties are to cognitive faculties), or at the very least that we are justified in doing so in the absence of any other evidence. If this is so, transcendental idealism is trivial even if true. Once we have formed the belief that God exists in order to justify the belief that P(R) is high, we have good reason to believe that P(F) is high as well. If this is the case, then the appearances in our understandings would be (nearly) perfect reflections of the things in themselves they represent.

Kant can stave off the collapse of his system into skepticism only by making concessions allowing for increases in the possible scope of human knowledge, and once he goes far enough to succeed he ceases to be a proponent of transcendental idealism at all. Belief in R can only be justified by a sufficient level of knowledge about things in themselves, and all such knowledge lies outside the bounds of transcendental idealism by definition. The theory’s epistemic

constraints are too narrow to support it, and any broadening renders it unrecognizable. Only by learning about the world as it truly is can we validate those faculties that allow us to learn in the first place, and thus it is theories that allow for this kind of knowledge that we must concern ourselves with.

Endnotes

¹ That is, placing bounds upon it which limit it profoundly but not so much as to constitute skepticism. This seems to me to be a key goal of the Kantian project.

² Mirza, Omar. "A User's Guide to the Evolutionary Argument Against Naturalism." *Philosophical Studies* 141 (2008): 127-28.

³ Plantinga, Alvin. "An Evolutionary Argument Against Naturalism." *Philosophy of Religion: An Anthology, Fifth Edition*. Louis P. Pojman and Michael Rea, eds. Belmont, CA: Wadsworth, 2003. 491.

⁴ Mirza, Omar. "A User's Guide to the Evolutionary Argument Against Naturalism." *Philosophical Studies* 141 (2008): 131.

⁵ Lewis, C.S. *Miracles*. New York, NY: HarperCollins, 2001. 21-22.

⁶ Schopenhauer, Arthur. "Sketch of a History of the Doctrine of the Ideal and the Real." *Parerga and Paralipomena*, Vol. 1. New York, NY: Oxford University Press, 2000. 11.

⁷ Ibid.

⁸ Kant, Immanuel. *Prolegomena to Any Future Metaphysics*. 2nd ed. Trans. James W. Ellington. Indianapolis, IN: Hackett, 2001. §52c B341-42.

⁹ Save for certain very general principles that may govern them. See: Zuckert, Rachel. *Kant on Beauty and Biology*. New York, NY: Cambridge University Press, 2007. 27.

¹⁰ Kant, Immanuel. *Critique of Pure Reason*. Rev. ed. Trans. J.M.D. Meiklejohn. New York, NY: Colonial Press, 1900. 89-90.

¹¹ Kant, Immanuel. *Prolegomena to Any Future Metaphysics*. 2nd ed. Trans. James W. Ellington. Indianapolis, IN: Hackett, 2001. §19 B298.

¹² Ibid., §57 B354-55.

¹³ Ibid., §58 B359.

¹⁴ Ibid., §57 B355, §58 B358-59.

¹⁵ Ibid., §58 B359. Emphasis added.