

The Rational Basis of Self-Trust

Karl Schafer (*Work in progress - do not cite without permission*)

March 31, 2011

“Have courage to make use of your *own* understanding! is thus the motto of enlightenment”¹

Kant’s famous “motto of the enlightenment” tells us to make use of our own faculties in deciding what to believe and what to do. Or, to put it more succinctly, it tells us to *think for ourselves*. On this point, it appears to me that Kant was fundamentally correct. Like Kant, I believe that the demand to think for oneself is one of the primary rational obligations we possess in both the theoretical and the practical spheres. And, again like Kant, I believe that the recognition of this requirement is a debt we owe, in large part, to the Enlightenment, even if our own recognition of this requirement is often far from perfect.

In the following I want to explore several related ways of arguing for the principle Kant gives voice to in his motto. In doing so, I hope to make this principle more plausible than it might at first seem. And I hope to do so in a manner that indicates why Kant’s “motto” deserves to be regarded not as a historical curiosity - but rather as a central component of a satisfying anti-skeptical account of epistemic rationality.²

¹Kant, “What is Enlightenment”, 8:35

²Principles of this general sort, at least in the epistemic case, are endorsed by a number of contemporary epistemologists. For example, see Richard Foley’s *Intellectual Trust in Oneself and Others* (Cambridge University Press, 2001) and Michael Huemer’s “Compassionate Phenomenological Conservatism” (*Philosophy and Phenomenological Research* (2007) 74:1). The application of this principle to practical reasoning has, to my knowledge, received much less attention.

My view is, in some respects, quite close to Huemer’s. But it does differ in a number of significant respects. First, I conceive of self-trust in terms of an *a priori* entitlement to believe that your faculties are reliable. Second, unlike Huemer, I don’t necessarily think of the relevant sort of self-trust in phenomenological terms. (This is connected with the sort of internalism I’m attracted to, which spells out “access” in terms informational or functional access as opposed to accessibility to phenomenal consciousness.) And, third, I

One of my main aims in the following is to discuss the manner in which an entitlement to trust one's own faculties is best understood. The most popular contemporary views of this general sort are forms of what, following the work of Jim Pryor, has come to be called Dogmatism - views according to which we should trust our faculties, whether or not we are independently justified in believing them to be reliable.³ But views of this general sort face a number of serious objections. Thus, I believe that our basic entitlement to trust our own faculties is best understood in terms of a basic entitlement to believe that our faculties are generally reliable. Views of this general sort have been proposed by number of recent epistemologists.⁴ But these authors' discussion of this sort of entitlement often leaves it unclear how it could be the case that we have an *a priori* entitlement to believe a substantive, deeply contingent fact of this sort. Is the existence of this entitlement simply a brute fact about epistemic rationality? Or is there some further explanation of why this entitlement holds? In the following discussion, I will offer a novel explanation of this sort of entitlement exists. In doing so, I will argue that given a plausible characterization of the obligations implicit in the very idea of epistemic rationality, we can derive from these obligations an *a priori* entitlement to believe that our faculties are reliable. Thus, there need not be anything mysterious in the claim that epistemic rationality demands of us that we believe *a priori* that our faculties are generally reliable.

1. Thinking for Oneself

In order to understand Kant's motto we need to first understand what it means to think for oneself. In particular, if we accept an obligation to think for ourselves, I believe that we should also accept that we possess the following basic rational entitlement to trust our faculties:⁵

(Self-Trust) Whenever you lack any positive reason to be suspicious of them, you are entitled to trust your own faculties.⁶

give an alternative explanation of why a principle like self-trust is true.

³See his "The Skeptic and the Dogmatist", *Nous* 34 (2000), 517-49

⁴See in particular the work of Crispin Wright and Roger White discussed below.

⁵To be clear, I am not claiming that this principle is all that rationality requires of us.

⁶This principle is meant to capture the demands that epistemic rationality places upon us as believers - and not the demands that practical rationality or morality places upon us as agents. Thus, the scope of this entitlement extends only to how we form beliefs, where this is evaluated from the point of view of epistemic rationality and where this does not include any further practical questions of what actions to perform or what intentions to form.

In order to understand Self-Trust, we need to understand two things: (i) what a positive reason to be suspicious of some faculty is, and (ii) what it is to trust a faculty.⁷

Let's take these in turn, focusing on the case of faculties for forming beliefs. As I understand it, you have positive reason to be suspicious of some faculty just in case some subset of your overall mental state is such that this subset would justify believing this faculty to be unreliable.⁸

So, for instance, you would have positive reason to be suspicious of one of your perceptual experiences, E, if you (justifiedly) believed E to be non-veridical. And you would have positive reason to be suspicious of E if you had other perceptual experiences whose contents conflicted with the content of E. For in this case, these other experiences would give you evidence that the world is not as E makes it seem it is. But if E was the only relevant mental state of yours, you would not have positive reason to be suspicious of it.

The nature of defeat is relatively straightforward in the perceptual case, and other cases like it, which will be my focus in the following discussion. But it is important to note that other cases, like the case of inference, are more complex than this. Suppose for example, that Jim believes P. And suppose that Q follows via a simple deductive rule from P. But also suppose that it seems obvious to Jim that NOT Q follows from P instead. In this case, it is *possible* that Jim has a defeater for both belief in Q *and* belief in NOT Q. After all, since Q follows from P via a simple deductive rule, Jim's belief that P justifies him in believing that Q, and so serves as a defeater (in the sense just defined) with respect to belief that NOT Q. But, on the other hand, given Self-Trust, the fact that it seems to Jim that NOT Q follows from P - together with his belief that P - *might* all other things being equal justify Jim in believing that NOT Q, and so might serve as a defeater with respect to belief that Q. In case, such as this, while Jim is obviously not justified in forming the belief that NOT Q, he may also not be justified in believing Q, given the basic conflict

⁷One might wonder whether this captures the full scope of the demand to think for yourself. Can we, after all, reduce trust in ourselves to trust in our faculties? Or are there other forms self-trust might take? It seems to me that if a belief is not the product of one or more of our faculties, it should not be regarded as the product of oneself in the sense that is relevant to self-trust. If so, then the above principle does capture the demand to think for oneself. But, whether or not this is the case, the cases that interest me most are clear cases in which what is at issue is whether or not we should trust one of our faculties for forming beliefs.

⁸I mean this principle to be compatible with both the view that the justifying mental state constitutes evidence of unreliability and the view that it simply gives one access to such evidence. Nothing in what follows requires that I take a stance on the question of whether evidence is best understood in terms of mental states or in terms of propositions or facts.

between the correct rules of deductive inference and the rules that seem obvious to him.⁹

Since my focus here will be on simpler cases, like the case of perception, I will for the most part set these sorts of complications aside. But it is important to keep them in mind, when considering what I have to say. Thus, let us turn to the second question noted above: i.e. what is it to trust a faculty?

Even if we focus on the case of faculties for forming beliefs, this can be understood in two distinct ways. The crucial issue in this regard is whether trusting a faculty in the relevant sense involves believing, implicitly or explicitly, that this faculty is reliable. Depending on how we answer this question, we will arrive at one of the following two characterizations of what it is to trust a faculty F:

(Self-Trust without Belief): Trusting a faculty F is a matter of being disposed to form beliefs on the basis of the deliverances of F, where this does not require that one believe that F is reliable.

(Self-Trust as Belief): Trusting a faculty F is a matter of believing, implicitly or explicitly, that F is reliable.

If we understand the demand that we trust our own faculties along the lines of the second of these characterizations, then an essential aspect of Self-Trust will be an explicit or implicit belief in the reliability of one's faculties. Thus, if we understand Self-Trust in this manner, an attitude of self-trust will be rational only insofar as this belief is rational. So if we conceive of Self-Trust in this way we arrive at the following principle:

(Entitlement to Believe) If you lack any positive reason to be suspicious of a faculty, you are entitled to believe that it is generally reliable.

This, though, may well be thought to represent an unnecessarily strong form of Self-

⁹There is a great deal more to be said about the issue of defeat. I believe that the most satisfying characterization of when some mental state M1 defeats the ability of some second mental state M2 to justify some third mental state M3 is roughly the following: M1 defeats our justification to enter into M3 on the basis of M2 just in case M1 justifies some further mental state M4 that is rationally incompatible with M3. But giving a full account of what this involves requires us to specify the conditions under which two mental states are rationally incompatible with one another. Elsewhere I take up this project, beginning with the basic rational obligations I discuss later on in this essay. But for our purposes here, we can work with the simpler characterization of defeat given above. (A characterization that can, I believe, be shown to be extensionally equivalent to the more complicated and informative characterization just given.)

Trust. For it may well be thought that, in the absence of a positive reason to be suspicious of them, we are entitled to form beliefs on the basis of the deliverances of our faculties, *whether or not we are justified in believing them to be reliable*. This is the sort of position that generally bears the label “Dogmatism” in the contemporary literature.¹⁰ And while a position of this general sort sits uncomfortably with the second way of understanding Self-Trust just laid out, it is wholly compatible with the first of these. Thus, if we conceive of Self-Trust in terms of Self-Trust without Belief, it may seem that there is no need to defend the conception of Self-Trust that Entitlement to Believe lays out.

Unfortunately for the Dogmatist, there are a number of objections to the idea that our entitlement to trust our faculties should be understood in this way. For example, one notable feature of Dogmatism is that it licenses coming to believe that one’s faculties are reliable by first relying on these faculties to form a series of particular “first-order” beliefs and then inferring from these particular beliefs, and one’s beliefs about the deliverances of one’s own faculties, the further result that these faculties are generally reliable. But if we allow for this sort of reasoning, then it seems we must also allow for a sort of epistemic “bootstrapping” that is viewed by many epistemologists as obviously irrational. In particular, on such a view, it appears that there will be nothing irrational about looking at one’s hand and reasoning as follows: “I have a hand, my perceptual experience makes it seem to me that there is a hand, therefore (in this respect) my perceptual experience is getting things exactly right.”¹¹

Such reasoning has seemed to many epistemologists to be entirely worthless. Personally, I’m not entirely convinced that this is the case. For example, while such arguments may not generate new propositional justification for their conclusions, they may put one in a position to form a doxastically justified belief in these conclusions, where one had not be

¹⁰It is important to stress here that there are a variety of more and less demanding ways of defining the essential elements in the Dogmatist’s position. Since my primary focus in the following arguments will be the existence of *a priori* entitlements to believe that our faculties are reliable, I have defined Dogmatism to highlight this issue. But there are other ways of characterizing what is essential to “Dogmatism”, not all of which are subject to all of the objections noted here.

For more on Dogmatism, see Pryor’s “The Skeptic and the Dogmatist”, as well as Tyler’s Burge’s “Content Preservation” (in *The Philosophical Review*, 1993: 457-488) and Chris Peacocke’s *The Realm of Reason* (Oxford: Oxford University Press, 2006).

¹¹See, e.g. Jonathan Vogel’s “Reliabilism Leveled” in *Journal of Philosophy* 97: 602-623 (2000) and Stewart Cohen’s “Basic Knowledge and the Problem of Easy Knowledge” in *Philosophy and Phenomenological Research* (2002) 65 (2): 309-329.

capable of doing so before.¹² But I am skeptical that reasoning of this sort can play the role that Dogmatists generally think of it as playing - and, in particular, I am skeptical that such reasoning be the source of new propositional justification for its conclusion.¹³ So even if there is something epistemically useful about such reasoning, this does not seem to me to remove the need to think of us as having some sort of initial *a priori* entitlement to believe that our faculties are reliable. Or, in other words, it does not seem to remove the need to accept some form of Entitlement to Believe as defined above.

Moreover, there are further problems for the Dogmatist's view that seem just as serious as the one just noted. For example, Dogmatism appears to be incompatible with the standard Bayesian of updating in terms of conditionalization.¹⁴ Thus, the acceptance of the Dogmatist position appears to require a significant revision in our understanding of partial belief.¹⁵ For these reasons, I am skeptical that we should understand the demand to trust our own faculties in the manner the Dogmatist suggests. Thus, whether or not we think of Self-Trust in terms of Self-Trust without Belief, I am inclined to the view that Self-Trust is only defensible insofar as something like Entitlement to Believe is true. Or, in other words, we should interpret Self-Trust as involving a defeasible *a priori* entitlement to believe that our faculties are generally reliable.¹⁶

¹²Compare Tim Willenken's discussion of dogmatism in "Moorean Responses to Skepticism: A Defense" forthcoming in *Philosophical Studies*.

¹³For one way of characterizing what is problematic about the idea that such reasoning might be a source of propositional justification, see Michael Titelbaum's "Tell me you love me: bootstrapping, externalism, and no-lose epistemology" in *Philosophical Studies* (2010) 149:119-134.

¹⁴For examples of this objection see: Roger White's "Problems for Dogmatism." *Philosophical Studies* 131 (2006) and Stephen Schiffer's "Skepticism and the Vagaries of Justified Belief," *Philosophical Studies* 119 (2004).

¹⁵In addition, there are significant problems for the Dogmatist when it comes to distinguishing the Moorean arguments he accepts as rational from other, structurally similar, arguments that seem obviously circular and so irrational. For important discussion of these issues, see e.g.: Crispin Wright's "Warrant for Nothing (and Foundations for Free?)" *Aristotelean Society Supplement* Vol 78 (2004), Jessica Brown's "Doubt, Circularity and the Moorean Response to the Sceptic," *Philosophical Perspectives* 19 (2005), and Martin Davies' forthcoming Hempel Lectures.

¹⁶This having been said, most of the argument below is independent of this point. Thus, even a Dogmatist might find much in the following discussion of why we are entitled to trust our faculties that would help to support and explain his conception of such self trust.

In addition, as I discuss below, the main argument of this paper is also compatible with forms of coherentism that allow for the existence of substantive *a priori* constraints on the shape that a "coherent" set of well-justified beliefs can take. The differences between this sort of view and a view that accepts Entitlement to Believe largely relate to more general debates between coherentists and foundationalists that lie outside

This way of conceiving of the demand that one think for oneself avoids the issues noted above for the Dogmatist position. For given the existence of this sort of *a priori* entitlement to believe that one's own faculties are generally reliable, the problematic forms of argument that the Dogmatist endorses are no longer relevant. On this sort of view, one can never acquire new propositional justification for the belief that one's faculties are reliable in the manner the Moorean describes, precisely because the level of trust one is entitled to have in one's faculties is bounded by one's prior entitlement to believe that these faculties are reliable.¹⁷

Of course, understanding Self-Trust in this way brings with it its own problematic commitments. In particular, if we accept the existence of an *a priori* entitlement of this sort, we need to explain why it is reasonable to think that we can be entitled *a priori* to a substantive belief of this sort about our mind and its relationship to the external world. One of my primary aims in the following discussion is to provide an explanation of just this sort. Thus, if my discussion below succeeds, we should not reject this option on these sorts of grounds. Thus, in the following I will understand Self-Trust here in second way noted above. But most of the discussion to follow will apply equally well (modulo the obvious changes) to a Dogmatist interpretation of Self-Trust. So readers with Dogmatist sympathies should find it easy to adapt this discussion to fit their views.

However it is understood, it should be clear that Self-Trust is an extremely general and powerful epistemic principle. For Self-Trust claims that for any of our faculties, F, we should trust the output of F, unless we are in a mental state that provides us with

of the scope of this paper. So I will not focus on them here.

¹⁷Thus, not only is it impossible to "bootstrap" oneself into a rational belief in the reliability of one's faculties on this view, it is also impossible to "bootstrap" oneself from belief in the reliability of one's faculties to rational belief in their "super-reliability" in the problematic manner Jonathan Weisberg describes in his "Bootstrapping in General" (forthcoming in *Philosophy and Phenomenological Research*). In both cases, the degree to which one is entitled to believe that one's faculties are reliable after their use will be a straightforward - if sometimes complicated - product of one's prior confidence in their reliability, plus their results, in much the manner Weisberg himself describes.

Having said this, it is worth noting that while these arguments cannot generate new propositional justification on the sort of view I endorse, they may be able to play other epistemically significant functions in some cases - such as placing the subject in a position to recognize his *a priori* entitlement to believe that his faculties are reliable.

evidence that F is unreliable.¹⁸ So if Self-Trust is true, a long list of other more local epistemic principles follow from it. For example, reading these principles in the manner just suggested:

(Perceptual Self-Trust) If you lack any positive reason to be suspicious of your perceptual experiences, you are entitled to believe that they are generally reliable.

(Mathematical Self-Trust) If you lack any positive reason to be suspicious of your mathematical intuitions, you are entitled to believe that they are generally reliable.

(Moral Self-Trust) If you lack any positive reason to be suspicious of your moral intuitions, you are entitled to believe that they are generally reliable.¹⁹

And so on. Thus, a principle like Self-Trust provides a very general basis for thinking about epistemological questions of very different sorts. In this way, Self-Trust is similar to traditional forms of externalism such as reliabilism. One of the attractions of reliabilism is that it provides us with a single, elegant standard for deciding when we should rely on a belief-forming faculty: namely, that it is rational to rely on a faculty just in case it is reliable (under certain specified conditions).

Self-Trust similarly provides us with a principle that specifies in simple and elegant terms when it is rational to rely on any belief-forming faculty. But unlike the reliabilist's standard, the standard set out by Self-Trust is at least potentially internalist in character. Given Self-Trust, whether one should rely on some faculty F depends only on whether one has positive reason to be suspicious of F. And so, if we are internalists about what it is to have such a reason, and we accept Self-Trust, then we will also be internalists about when we rely on trust our faculties. Thus, Self-Trust, unlike reliabilism, is fully compatible with an internalist account of these issues.²⁰

¹⁸There is, of course, the question here of just which faculties should count as "ours" in the relevant sense. Given the argument below, the faculties to which Self-Trust applies will be just those that make a contribution to the way things seem to us from a first-person point of view. But for the time being we can postpone discuss of this question.

¹⁹To be clear here, I don't mean to put any serious weight on this way of craving up our different faculties. Since Entitlement to Believe treats us as having an *a priori* entitlement to believe that *all* our faculties are reliable, there's no need to settle how we should individuate our faculties in interpreting this principle. The relevant distinctions between the faculties we are entitled to trust and those we are not entitled to trust will, on this view, simply fall out of the defeating evidence we possess.

²⁰Nonetheless, it is important to stress that Self-Trust is also compatible with a much more externalist

In particular, suppose we conceive of internalism in terms of the following characteristic internalist claim:

(Internalism) The nature of your rational entitlements supervenes on facts to which you have internal access.²¹

Given this characterization, Self-Trust will be compatible with Internalism just in case whether or not someone has positive reason to be suspicious of some faculty is something that supervenes on facts to which they have access in the relevant sense. So Self-Trust is compatible with the general internalist idea that whether one should trust some faculty supervenes on facts to which one has access, provided we accept an internalist conception of what it is to have positive reason to be suspicious of a faculty.

In this way, Self-Trust provides us with a way of giving an internalist account of what one should believe that matches the elegance of externalist accounts like reliablism. But it does so only via recommending that we trust all of our faculties for belief formation *without exception* so long as we lack positive reason to doubt them. And this may seem to have some very counterintuitive epistemic consequences.²²

Perhaps most importantly, there is no guarantee that beliefs formed in the manner Self-Trust recommends will, even in general, be more likely to be true than false. Here the contrast between a principle like Self-Trust and reliablist accounts of what we may believe is very clear. For one of the main attractions of reliablism is that it is built around

conception of reasons to be suspicious of some faculty.

²¹The label “internalism” is, of course, applied to a wide range of different views in contemporary epistemology, not all of which are precisifications of this formula. So it is important to stress that here, and in the following, that what I have in mind by “internalism” is a view of this general sort. Whether or not other forms of internalism are compatible with Self-Trust is further issue that I will not address here.

In addition, the notion of access at work here can be spelled out in a number of different ways - e.g. in terms of phenomenal consciousness or in terms of informational access of some sort. Since I don't want to take a stand here on the proper way of developing the internalist's position, I want to remain content with the highly schematic conception of internalism above. All that is important to me here is that Self-Trust will be compatible with Internalism, provided that whether or not someone has position reason to be suspicious of some faculty is something that they can have “internal access” to whatever sense one takes to be characteristic of Internalism.

²²In particular, the position developed here is compatible with a stronger form of internalism than many similar positions. Compare, for example, the discussion of the New Evil Demon Problem in David Enoch and Josh Schlechter's “How are Basic Belief-Forming Methods Justified?”, forthcoming in *Philosophy and Phenomenological Research*.

a connection of this sort. But even someone who rejects reliabilism might well believe that there must be *some* connection between any true epistemic principle and the truth of the beliefs formed in accordance with it. For it can seem that in order for a principle to count as an *epistemic* principle in the first place, it must have some such connection with the truth.

Does Self-Trust involve any connection of this sort? Well, Self-Trust simply tells us to trust our faculties - so long as we lack any positive reason to be suspicious of them. And of course there is no guarantee that our faculties will be reliable. So there is certainly no guarantee that Self-Trust will lead us to form true beliefs. In fact, abstracting away from our trust in our own faculties, we seem to have no reason to think that these faculties are more reliable than not. So we have no reason, independent of our own self-trust, to think that forming beliefs in the manner Self-Trust recommends will be more likely to produce true beliefs than false ones.

In this sense, there is no connection between following Self-Trust and being likely to form true beliefs beyond the obvious claim that if our faculties *are* generally reliable, then forming beliefs in the manner Self-Trust recommends will tend to produce true beliefs. But of course the same will be true of many methods for forming beliefs that are obviously irrational. So whatever it is that explains the rationality of Self-Trust, it cannot be the fact that compliance with Self-Trust has any sort of special connection with the truth.

To some, this alone will render a principle like Self-Trust unsuitable to play a foundational role in the proper account of epistemic rationality. But it seems to me that to demand a stronger connection than this between the principles of epistemic rationality and the truth is to demand something that we have no right to expect. After all, there is no guarantee that we are not in circumstances in which our faculties are systematically unreliable. And yet if we (unbeknownst to us) happened to be in a situation of this sort, it seems to me that we would be entitled to continue to form beliefs via trusting our own faculties just as we actually do.²³ In which case, it seems that we cannot expect the principles that govern how we should form beliefs to have a stronger connection with the truth than Self-Trust does.²⁴

²³Here I am, of course, simply giving voice to my internalist sympathies. Once again, my goal in the present discussion is not to engage in a serious discussion of the respective merits of different forms of internalism and externalism. All I want to do is to stress the compatibility of Self-Trust with an internalist conception of them. For related discussion see Stewart Cohen's discussion of the "New Evil Demon Problem" in his "Justification and the Truth" *Philosophical Studies* 46 (1984) 279-295.

²⁴Of course, we might try to buttress this connection in various ways. (See, for example, Ralph Wedg-

Simply saying this, though, is obviously insufficient here. For it is not enough to reject the conception of epistemology that brings with it the demand that facts about epistemic rationality should be explained solely in virtue of the special relationship between epistemically rational methods and the truth. We also need to describe an alternative conception of epistemic rationality, explain why this way of conceiving of epistemic rationality is plausible, and show that it implies that Self-Trust is an epistemically rational way to form beliefs. It is to this task that I turn now.²⁵

3. Deriving Self-Trust from Our Basic Epistemic Obligations

With this in mind, let's return to the question of whether there is anything further we can say in defense of Self-Trust. I believe there is. In particular, I believe that it is possible to give an argument for Self-Trust that begins with the claim that we possess a basic epistemic obligation to form beliefs that meet certain criteria, and then argues that it is only possible for us to comply with this obligation in so far as we form beliefs in then manner that Self-Trust instructs.²⁶

wood's discussion of internalism in "Internalism Explained" (*Philosophy and Phenomenological Research* 65 (2002), 349-369.) I'm skeptical about the sorts of claims philosophers like Wedgwood rely on to make a connection between these sorts of principles and the truth. For more on this issue, see below.

²⁵In this essay, I will be concerned primarily with the implications that Self-Trust has for when we should or should trust our own faculties. But it is important to note that Self-Trust also has important implications for our relationship with the faculties of other individuals. After all, suppose I trust my own faculties in the manner Self-Trust recommends. Doing so implicitly commits me to believing that these faculties are at least generally reliable. But if I believe that my own faculties are generally reliable, how should I view the faculties of other individuals around me? Well, in so far as I regard the faculties of others as similar to mine in the relevant respects, I should regard their faculties as being just as reliable as my own. So, if I regard my own faculties as reliable, this commits me either to regard the faculties of others as reliable or to regard their faculties as different from my own in some relevant respect. And this, given our beliefs about the similarities between our faculties, will generally commit us to treating the faculties of others as reliable. For more on arguments of this sort, see Richard Foley's "Universal Intellectual Trust", *Episteme*, 2005, as well as my XXX.

²⁶There are a number of arguments in the literature that have similarities to the argument I will be sketching here, although all of them differ from it in important respects. For example, certain recent remarks of Crispin Wright and Paul Boghossian suggest a picture of epistemic rationality that has the same general flavor as the account I am developing here. (See Wright's "Warrant for Nothing", *Proceedings of the Aristotelian Society* Supp. Vol. 78 (2004): 167-212 and Boghossian's "Blind Rule-Following," in *Mind, Meaning and Knowledge: Essays for Crispin Wright*, ed. Annalisa Coliva (Oxford: Oxford University Press, 2007))

Moreover, David Enoch and Josh Schlechter's work on Reichenbachian arguments for the rationality of

In other words, the sort of argument I have in mind has the following structure:

- (1) Self-trust is required in order for us to A.²⁷
- (2) A'ing is rationally obligatory.
- (3) If A'ing is rationally obligatory, and self-trust is required for us to A, then trusting oneself in order to A is also rationally obligatory.
- (4) Therefore, trusting oneself in order to A is rationally obligatory.

This is an argument for the obligatoriness of self-trust. But if it is successful, it also supports the weaker conclusion that self-trust is permissible - or that we are entitled to trust our faculties - which is the conclusion of interest to us here.²⁸

But is it really true - as premise (3) of the argument above claims - that the necessary prerequisites of an obligatory activity are always obligatory?²⁹ A natural worry about this claim is that some activity A might be obligatory in general, but this obligation might be defeated by the fact that A'ing requires one to trust oneself, thereby making (3) false. For example, suppose we are obligated to keep our promises. Even if this true, when keeping a promise would require us to seriously hurt another human being, this obligation is defeated and we are in fact obligated to break our promise. And if this is the case, then our obligation to keep our promises does not translate into a general obligation (or permission) to do whatever is required to keep them.

This does not mean that the very idea of giving this sort of argument is confused.

basic belief-forming methods rely on considerations that share many structural similarities to the arguments I am giving here for Self-Trust - although the "pragmatic" characterization of these arguments that they provide is different than the characterization I give here in several important respects. (See their "How are Basic Belief-Forming Methods Justified?", forthcoming in *Philosophy and Phenomenological Research*.) Finally, Ralph Wedgwood is currently in the process of developing an account of these issues that shares some of the features of the account I am presenting here. Each of these accounts, though, takes on commitments that the argument I am offering here avoids. And none of these authors, I believe, would agree with the diagnosis I offer below of the basic lesson we should learn from an encounter with perceptual skepticism.

²⁷Arguably, this premise is stronger than we require in every case, for it may be the case that an argument of this sort may be run (at least for the permissibility of self-trust) beginning with the premise that self-trust is *as good a means* for successfully engaging in A as anything else. (This is the sort of argument Reichenbach attempts in the case of induction.)

²⁸Could we run an argument for the permissibility of self-trust beginning with the weaker premise that A'ing is only permissible? Perhaps, but this sort of argument raises some delicate issues about rational decision making under conditions of uncertainty, which seem to me clearer if we focus on arguments whose premises involves claims about what is obligatory. For more on this issue, see below.

²⁹Or, alternatively, are the necessary prerequisites of an obligatory activity always permissible?

Rather, it indicates that when giving this sort of argument the obligations we invoke had better not be open to these sorts of defeating considerations. Or, in other words, these obligations had better be indefeasible - or, at least, they'd better be indefeasible, given that the background conditions we're interested in obtain. So, for example, if we want to argue that we have an obligation to rely on our perceptual experiences in forming beliefs whenever we are unaware of any relevant defeaters, we must do so by appealing to an activity A that is indefeasibly obligatory under these conditions. For the necessary prerequisites for an activity that is *indefeasibly* obligatory (given some condition) will themselves be obligatory (given this condition). After all, if they were not, the fact that they were prerequisites of this activity (given this condition) would itself be a defeater for our obligation to engage in it.³⁰

Thus, if an argument of this sort for Self-Trust is to be successful, its second premise must be the following:

(2') Given the relevant background conditions, A'ing is *indefeasibly* rationally obligatory.

Thus, if there is a satisfactory argument of this sort for Self-Trust, there must be some activity of which the following two things are true: First, we can successfully engage in this activity only by trusting our faculties. And, second, this activity is rationally obligatory and this obligation is indefeasible under the relevant background conditions.³¹

(1) *Truth or Accuracy*: Is it possible to find an activity A that meets these demands? The most obvious candidate for such an activity is the activity of forming a set of beliefs that are (at least in general) true. But I do not believe that this activity, on its own, provides a plausible basis for Self-Trust.

To see why, we need to first note that there are two quite different ways we might understand the role of truth or accuracy in grounding facts about how we should form

³⁰There is an interesting objection to these claims that relates to the nature of rational decision under conditions of uncertainty. Since discussing this line of objection requires dealing with some rather complicated issues in this area, I postpone it to an appendix to this essay. Those who are interested should turn there now.

³¹In searching for an activity that is capable of playing this role, I am in effect hoping to show that it is possible to give what Huemer describes as a "meta-justification" for principles like Self-Trust. Although it is important to stress that it will only be possible for me to do because I leave behind the aim of giving such a justification solely in terms of the (likely) reliability of trusting oneself.

beliefs. On the one hand, we might say that truth is relevant here because we are obligated to only believe what *seems* to us to be true. Or, on the other, we might claim that truth is relevant here because we are obligated to form beliefs that (at least in general) *are* actually true.

Of these two, only the first is remotely plausible as a rational obligation. After all, to say that we are rationally obligated to form a set of beliefs that are (at least in general) true is just to say that the overall accuracy of our set of beliefs is a condition on our being rational. But however we understand the demands of rationality, it should be possible for someone to form beliefs in a perfectly rational manner and yet, through sheer bad luck, end up with a set of beliefs that are largely false. In other words, on no plausible conception of rationality is being rational a guarantee that one will get things right, even in general.

This, it is important to stress, is true even with respect to externalist theories of epistemic rationality such as reliabilism. For while these theories do demand that one form beliefs via methods that (say) are reliable under certain circumstances, reliability does not *guarantee* that one will get things right, even given these conditions. So, even from an externalist perspective, the view that we are epistemically obligated to believe the truth should seem unattractive.

But what of a more modest externalist connection between epistemic rationality and the truth? For example, what of the idea, common to all forms of reliabilism, that we are epistemically obligated to form beliefs via methods that are generally reliable? Certainly there is something attractive about this idea - and I do believe that reliability is a necessary condition on achieving certain positive epistemic statuses such as knowledge.³² But in the end, I do not think it is plausible that reliability is a necessity condition on epistemic rationality.

My reasons for rejecting this possibility are simply that it seems to me clear that there are possible situations in which individuals are rationally obligated to form beliefs via methods that fail to be reliable in whatever sense one takes to be relevant. For example, it seems to me clear that individuals in an Evil Demon World are not rationally obligated to resist trusting their senses, so long as they are unaware that their senses are in fact unreliable. Rather, it seems to me clear that these individuals have precisely the same epistemic obligations that we do in the actual world.³³

³²For more on this, see my XXX.

³³Of course, there is lengthy literature concerning how externalist might be modified to deal with these sorts of cases. For example, the externalist might modify his view so that it focus upon, not the reliability

One might deny this by giving a thoroughly externalist account of our epistemic obligations. That is, one might claim that in such worlds we are not obligated to trust our senses simply because our senses are systematically misleading, whether or not we are aware of this fact. But while I have a good deal of sympathy with thoroughly externalist accounts of some epistemic concepts (for example, of knowledge), externalism to my mind has little plausibility with respect to what we are obligated to believe. For it seems to be basic to the notion of obligation that what we are obligated to believe supervenes on features of our situation to which we have access in some intuitive, if vague, sense. And this is enough to rule out an externalist account of our epistemic obligations.³⁴

For these reasons, it seems to me that the truth-oriented epistemic obligations we possess should not be understood in externalist terms. Rather, if there is an important epistemic obligation here, it must involve, not an obligation to believe the truth or an obligation to rely on reliable methods, but rather some sort of obligation that relates to what *seems* true to one.

It is easy to imagine several possible obligations of this sort. For example, we might suppose that we are obligated to believe P whenever it seems to us that P.³⁵ Or, we might accept the somewhat weaker principle that we are obligated to believe only what seems to us to be true.³⁶ Finally, we might accept the even weaker principle that we are obligated not to believe NOT P when (all things considered) it seems to us that P.³⁷

Unfortunately, it seems unlikely that any of these principles will provide the basis for an interesting argument for Self-Trust on their own. The problem here is simple. If we

of a method in the world in which that method is used, but instead the reliability of this method in the *actual world*. There are a number of issues with such views - such as whether they are faithful to the original motivations for reliabilism. But the most basic problem with such views remains, to my mind, the fact that they are insufficiently responsive to the intuitions that motivate internalism. After all, while such views with capture these intuitions insofar as we are concerned with the question of which beliefs would be rational in some scenario considered *counterfactually*, they will not capture these intuitions insofar as we are concerned with the question of which beliefs would be rational if some scenario were actual (i.e. if that scenario is considered *counteractually*). And any internalist worth his salt will have the same intuitions about either of these sorts of cases.

³⁴For more on how I think internalism and externalism should be understood see my XXX. Amongst other things, that essay argues that any satisfying account of the nature of epistemic evaluation will have to make room for epistemic concepts of *both* an externalist and an internalist character.

³⁵Compare Huemer's principle of Phenomenological Conservatism.

³⁶Compare certain forms of contemporary Evidentialism.

³⁷To be clear, there are issues with all of these principles. But since my argument below does not rely on any of them in particular, I will leave these issues aside here.

take one of the weaker principles of this sort as the basis of our argument - such as the last principle just noted - then this principle will be too weak to support an argument for Self-Trust. For this last principle only demands of us that *not* form beliefs that are contrary to the way things seem. And this, of course, is compatible with a total suspension of empirical belief of the sort the skeptic recommends.

On the other hand, if we choose one of the stronger principles just noted as the basis of our argument, the result will be an argument whose premises are too close to its conclusion for the argument to have much interest. For surely anyone who is skeptic about Self-Trust will also be skeptical that, say, we are obligated to believe P whenever it seems to us that P.

(2) *Power*: For these reasons, it seems unlikely that our obligations to believe the apparent truth will provide us with the basis of an interesting argument for Self-Trust. As will become clear in a moment, I believe that it is only possible to give an account of epistemic rationality that delivers this result if we expand our understanding of epistemic rationality to involve more than these sorts of truth-related obligations. This may be thought to involve an important concession to the skeptic. For if we limit ourselves to a conception of epistemic rationality that focuses only on what seems to us to be true, the skeptic is entirely correct to insist that these sorts of considerations give us no non-question-begging reason to trust our own faculties - as opposed to, say, simply suspending judgment about the world around us.

Fortunately for our anti-skeptical ambitions, our intuitive, everyday understanding of epistemic rationality extends beyond these truth-related obligations. What we are concerned with in forming beliefs is not just that these beliefs are true, but also that these beliefs possess a high degree of explanatory and predictive power. Thus, there is an important epistemic *desideratum* in addition to those that are connected directly with the truth. Namely, we would like our beliefs to provide us with predictions and explanations to the greatest possible degree. Following Hartry Field, we may call this desideratum, “power”.³⁸

Now, it might be thought that this obligation should be characterized in terms of an obligation to form a powerful set of beliefs. But according to most conceptions of explanatory power, a set of beliefs will only possess explanatory power insofar as these

³⁸For a nice exposition of this idea see Hartry Field’s “A Prioricity as an Evaluative Notion” in P. Boghossian and C. Peacocke (eds.) *New Essays on the A Priori* (Oxford: Oxford University Press 2000). Also see Alvin Goldman’s *Epistemology and Cognition* (Cambridge: Harvard University Press (1988)).

beliefs are true - or at least approximately true.³⁹ Given this conception of power, if we were obligated to form a powerful set of beliefs, we would also be obligated to form a set of beliefs that is at least approximately true. But, for the reasons discussed above, we are not rationally obligated to form a set of beliefs that are true, even approximately. So this cannot be the correct way of interpreting the role of power in epistemic rationality.

To avoid this problem, we need to think of this obligation in terms of an obligation to form a set of beliefs that, if true, would be powerful. Or, in other words, we need to conceive of this as an obligation to form a *potentially powerful* set of beliefs. Combining this with the comments made above, we arrive at a conception of epistemic rationality that demands that we form a set of beliefs that both (i) satisfy our obligations relating to apparent truth and (ii) that, if true, would be powerful. Of course, the obligation to form beliefs with both these characteristics is defeasible - for we can learn that it is impossible for us to form a set of beliefs with them. But this does not endanger our argument here, for learning this would also defeat the intuitive obligation to trust our faculties that we are trying to defend.

I think that this way of thinking of epistemic rationality corresponds very closely to our everyday understanding of it. For while we might say that truth (or at least “apparent truth”) is the first “aim” of belief formation, we generally have little interest in forming true beliefs where these do nothing to help us to explain and predict the events that matter to us. Thus, *neither* forming powerful beliefs *nor* forming true beliefs has much epistemic interest on its own for us.

I hope very few will seriously doubt that we possess *some sort of* rational obligation to form a potentially powerful set of beliefs. Still, many may wonder whether this obligation is a part of *epistemic* rationality as opposed to rationality in some broader sense. In this regard, it is very important to remember that the obligation we are considering is an obligation to *form beliefs* or *make judgments* in ways that allow one to form a potentially powerful set of beliefs. Thus, this obligation does not have any implications for the intentional actions we ought to perform. And, in particular, it does not demand of us that we perform *intentional actions* that will lead to the formation of such a set of beliefs. So the

³⁹This view of explanatory power has been widely accepted at least since Hempel. But it is far from uncontroversial, since there are many cases of apparently successful explanations which involve *explanans* that are not strictly speaking true. But even if we do not think that successful explanation requires that the explanans be true, we must - it seems - accept that there is some sense in which the *explanans* “approximates” the relevant truths, which is sufficient to raise the issue that concerns us here.

obligation that I am discussing here does not demand of one that one go out into the world and gather evidence about the nature of things. All that it demands is that one forms beliefs in a manner that, absent the presence of defeating considerations, and taking one's choice of actions for granted, allows one to form a potentially powerful set of beliefs.⁴⁰

Consider, for example, an individual who fails to form a powerful set of beliefs because he spends his entire life engaged in some deeply uninformative pastime, such as watching Teletubbies. Although such an individual would fail to form a powerful set of beliefs, it is not the case that he would fail to comply with the obligation we are considering here. For this obligation only concerns how the individual forms beliefs or makes judgments, and it is not this - but rather his refusal to perform certain actions - that prevents this individual from forming a reasonably powerful set of beliefs. In fact, given his choice of actions, there is no non-arbitrary way for this individual to form a reasonably powerful set of beliefs. Thus, given the rational obligations discussed below, there is no rational way for him to form a more powerful set of belief without going out and gathering more evidence. Thus, given his choice of actions, he is in fact complying with the obligation to form a powerful set of beliefs to the maximum degree that is possible, given the other rational obligations relevant here.

In this way, if there is a failure of rationality present in cases like this, it is failure of practical rationality - in particular, a failure to take the value of knowledge seriously enough from a practical point of view. But while such failures are obviously of great importance, they are not my subject here - for the obligations I am discussing here are narrowly epistemic insofar as they concern only how we should form beliefs, holding our choice of actions fixed - and not how we should *act* in light of our epistemic aims.

When it is restricted to how we form beliefs in this way, the obligation to form a potentially powerful set of beliefs seems to me to be central to our intuitive understanding of epistemic rationality. Insofar as I have an intuitive grasp on what epistemic rationality involves, forming a set of beliefs with explanatory and predictive power is every bit as much a part of epistemic rationality as is forming a set of beliefs that are accurate. In other words, what is central to our ordinary conception of epistemic rationality is not just a bare concern for truth and falsity as such, but rather a concern for the sort of understanding of the world that involves not just the formation of true beliefs, but also the formation of powerful sets of beliefs as well.

⁴⁰For related discussion of the distinction between epistemic and practical rationality, see Pamela Hieronymi's "The Wrong Kind of Reason" in *The Journal of Philosophy* 102:9 (2005): 437-457.

Evidence for this way of characterizing our everyday understanding of epistemic rationality is, I think, provided by the depth of our pre-theoretical commitment to methods of reasoning such as induction and inference to the best explanation. Generations of philosophers have attempted to justify these forms of reasoning by arguing that, say, there is some sort of metaphysical connection between simplicity and truth.⁴¹ But these attempts have proved notoriously problematic. So while it may someday prove possible to derive a connection of this sort, it seems implausible to think that our *everyday* acceptance of these methods as epistemically rational is motivated by a pre-theoretical acceptance of the thought that the truth *must* conform to these methods. Far more plausible to my mind, is the suggestion that we pre-theoretically take these methods to be epistemically rational because the acceptance of such methods appears to be required if we are going to be able to form a potentially powerful set of beliefs. In this way, the *deep, pre-theoretical* acceptance of these methods as epistemically rational provides us with an important source *prima facie* evidence that our ordinary conception of rationality embodies a concern for power as well as truth.⁴²

Of course, one may always define “epistemic” rationality so that it is concerned only with forming true (or seemingly true) beliefs. But the notion of epistemic rationality that focuses on truth alone (as opposed to truth and power) seems to me very far removed from our ordinary understanding of what epistemic rationality involves. So it is unclear, to say the least, why we should be concerned by the results we arrive at beginning with this artificial and restrictive conception of epistemic rationality.

For these reasons I believe that we should accept an epistemic obligation to form a potentially powerful set of beliefs. But does it follow from this that we must accept Self-Trust? Consider the obligation to form a set of beliefs that are potentially powerful. If we are going to comply with this obligation, it is obvious that we must trust some of our belief-forming faculties. For it is simply impossible to form a potentially powerful set of beliefs,

⁴¹Perhaps the most famous historical explanation of this line of thought may be found in Leibniz’s defense of the Principle of Sufficient Reason. For a recent attempt to revive a more moderate version of it, see Chris Peacocke’s *The Realm of Reason*.

⁴²Similar observations may be drawn from throughout the history of philosophy, where it is easy to find philosophers who stress the epistemic significance of forming a set of beliefs that are not just true, but which also represent a rich and systematic understanding of the nature of things. For example, compare the distinction (in Descartes and many others) between mere *cognitio* and true *scientia* and Kant’s conception of experience as discussed below. (Of course, each of these cases raise interpretative issues that there is no space to discuss here.)

without forming a set of beliefs that are connected together by a network of deductive and inductive inferences of various sorts.⁴³ Thus, it is simply impossible to form a potentially powerful set of beliefs without trusting our reasoning faculties, at least in some cases.⁴⁴ And similarly, it is simply impossible to form any complicated set of beliefs of this sort without relying upon our faculties for the retention of belief. So it is impossible to form a set of beliefs that is even potentially powerful without trusting our memory, at least in some cases.

For these reasons, it follows directly from the obligation to form a potentially powerful set of beliefs, that we must trust some of our faculties, at least some of the time. This takes us some of the way to Self-Trust, but obviously not all of the way there. For we might still choose to trust some of our faculties for empirical belief formation, but not others. And we might choose to trust these faculties only some of the time.

To close this gap, we need to combine our obligation to form a potentially powerful set of beliefs with an obligation not to make arbitrary distinctions. In particular, suppose that I have no positive reason to be suspicious of any of my faculties. And suppose that all of these faculties are capable of being a source of potentially powerful beliefs.⁴⁵ Then it seems that neither my obligation to believe the apparent truth nor my obligation to form a potentially powerful set of beliefs gives me any reason to distinguish between these faculties in any way. Thus, my basic rational obligations provide me with no basis for drawing a distinction between my faculties.⁴⁶ And so any such distinction would be epistemically

⁴³Here it is crucial that having a set of a potentially powerful beliefs involves more than simply having such beliefs - it also involves connecting these beliefs together with one another via certain inferences. After all, it is the inferential connections between beliefs that provide them with their explanatory and predictive power.

⁴⁴Or at least this will be the case, provided that our reasoning faculties are a potential source of a powerful set of beliefs. For some individuals - e.g. the counter-inductivist - this will not be the case. Thus, the argument to follow will not apply to these individuals. But individuals of this sort are so deeply irrational that it is unclear whether they possess an obligation to Self-Trust of the sort we are discussing at all. In any case, I'll set such cases aside for the purposes of the following discussion.

⁴⁵Here I abstract away from the possibility that some faculties may be a better source of potentially powerful beliefs than others. Whether this gives us any reason to put more faith in some of these faculties than others is a delicate question. But it does seem intuitive to me that when two faculties conflict with one another, we should tend to trust the faculty that is a more powerful source of potential information about the world. After all, something very like this seems to me implicit in degree to which we take our senses of sight and touch seriously as opposed to, say, our senses of taste and smell.

⁴⁶Of course, some of these faculties may be more reliable than others. And some, like my faculty for deductive reasoning, may even be necessarily truth-perserving. But the first of these facts is not relevant

arbitrary. Either I should trust all of these faculties - or none at all. And so, given the points just made, I should trust them all.⁴⁷

If this is right than we have the makings of an argument for Self-Trust. For we have an argument that shows that we can only comply with our basic epistemic obligations via doing what Self-Trust says we should. Or, more precisely:

(ST1) I can only form a set of beliefs with a reasonable degree of potential power, if I trust some of my faculties some of the time - such as my memory and my faculties for reasoning.

(ST2) If I have no positive reason to be suspicious of my faculties, forming a potentially powerful set of beliefs about the world is indefeasibly rationally obligatory.

(ST3) Therefore, if I have no positive reason to be suspicious of my faculties, trusting my memory and reasoning faculties some of the time is indefeasibly rational obligatory.

(ST4) If trust some of my faculties but not others, with no positive reason to be suspicious of any of them, I am making an epistemically arbitrary distinction.

(ST5) Avoiding epistemically arbitrary distinctions is indefeasibly rationally obligatory.

(ST6) Therefore, if I have no positive reason to be suspicious of my faculties, I am indefeasibly rationally obligated either to trust all of them or none of them.

(ST7) Therefore, if I have no positive reason to be suspicious of my faculties, I am indefeasibly rationally obligated to trust them all.

So, if we accept that we are obligated to form a set of beliefs that is potentially powerful, and accept that we should not draw arbitrary distinctions, then we will be in a position to derive from these obligations a further obligation to trust our faculties when no defeating considerations are present.

to my basic epistemic obligations as we are conceiving of them here. And while some of the faculties in question may be necessarily truth-perserving, others - like my faculties for inductive reasoning and my memory - are plainly not. So these sorts of considerations will not provide us with a distinction between these faculties that is relevant to our basic epistemic obligations as characterized above.

⁴⁷For a detailed discussion of the arbitrariness of restricting principles like Self-Trust, see Huemer's "Compassionate Phenomenological Conservatism". The Kantian argument I am giving here for Self-Trust is plainly connected with Huemer's argument there that the denial of his preferred version of self trust in self-defeating. Nonetheless, these are argument are quite different in their structure and their ambitions. In particular, I begin with a positive conception of epistemic rationality and argue that this conception demands Self-Trust, while Huemer is simply concerned to show that certain alternative positions are committed to epistemically arbitrary distinctions between our faculties.

Now, as we've already discussed to some degree, this obligation may be understood in several ways:

1. Following the Dogmatist's lead, we might understand it in terms of an obligation to believe that things are the way they seem to us, whether or not we have any positive beliefs in the reliability of our faculties. Although this is not my preferred way of understanding the obligation we have just derived, my disagreement with a position of this sort are relatively minor.
2. Alternatively, we might interpret this obligation in the second way sketched above. In other words, we might interpret it as implying that we have a defeasible *a priori* entitlement to believe that our faculties are reliable, at least insofar as they contribute to the way things seem to us. This, of course, is what Entitlement to Believe, as defined above, claims is the case. Insofar as I find the objections to Dogmatism noted above compelling, I believe it is better to interpret the conclusion to our argument in this manner, although once again, the reasons that lie behind this preference are not the focus of the arguments of this paper.
3. Finally, we might interpret the conclusion of this argument in broadly coherentist terms. The first two ways of interpreting this conclusion just noted conceive of epistemic rationality in broadly foundationalist terms. But we might also interpret the implications of this argument in terms of a holistic coherentism that demands of us only that we form an internally coherent and mutually supporting set of powerful beliefs that is compatible with certain basic rational constraints. On this interpretation, it would be an *a priori* constraint on the nature of any system of well-justified beliefs (formed in the absence of defeating conditions) that it involves self trust in one of the two forms discussed above.

Given the Kantian origins of the arguments I am exploring here, this last way of interpreting them may seem quite natural. But, in the end, I am not at all sure how much of substance separates a view of this sort from a view of the second sort which regards our *a priori* entitlement to believe ourselves reliable to be a matter of propositional, as opposed to doxastic, justification. For it is easy to flesh out such a view so that it delivers the same results concerning doxastic justification as this third sort of view does. Thus, the difference between these sorts of views may well come down to a difference of opinion regarding the -

at least somewhat technical - question of what we are *propositionally* justified in believing *a priori*.

In any case, which of these three views one prefers will depend on largely on one's views questions in epistemology that, to a large degree, are independent on the main arguments of this paper. So I do not want to take a strong stand on these issues here, except to note these three ways of developing the conclusion of these arguments. Nonetheless, for an epistemologist who is suspicious of both Dogmatism and coherentism, it will only be possible to respect the conclusion of this argument by accepting a principle like Entitlement to Believe. Thus, insofar as we are inclined to reject these epistemological positions, we can view this argument as showing us that such an entitlement does in fact follow from the basic epistemic obligations outlined above.

4. Epistemic Conclusions

In this way we can argue from two of our basic epistemic obligations to the existence of an epistemic obligation to form beliefs in the manner Self-Trust recommends. Of course, as I noted above, neither of these two epistemic obligations has any guaranteed connection with the truth.⁴⁸ Nonetheless, I believe that these two obligations are essential to our understanding of what it is to form beliefs in an epistemically rational fashion.

Two general lessons about epistemic rationality follow from these claims. First, while it is true that epistemic rationality does have a connection with the truth, this connection is not based on any general connection between being rational and being reliable. As I noted before, this should come as no great surprise to those already inclined to some form of epistemic internalism. But it does mean that the ultimate value of being rational cannot lie in rationality being a means towards reliability. Of course, if things are as we generally take them to be, then being rational will generally make us more reliable. But we have no reason, independent of our trust in our own faculties, to suppose that this connection between reliability and rationality in fact obtains. So our belief that rationality will generally make us more reliable cannot explain why we should try to act in an epistemically rational fashion.

Thus, the value of being epistemically rational cannot lie in any general connection between rationality and reliability. But if rationality is not necessarily a way of becoming

⁴⁸Naturally there are philosophers who do try to establish a connection of this sort. While their attempts seem to me to fail, they surely are of great interest. (For instance, see Christopher Peacocke's discussion of simplicity and truth in his *The Realm of Reason*.)

more reliable, why should we care about epistemic rationality in the manner we seem to? This is one of the most fundamental questions in epistemology. But the essence of an answer to it does not seem to me to be difficult to uncover - and it brings us to the second main lesson about epistemic rationality that I want to stress. The value of rationality lies, in the first instance, not in the fact that rationality is a way of producing an accurate system of beliefs - but instead in the fact that rationality is a way (*the* way, in fact) of producing a unified, coherent, and explanatorily powerful perspective on the world. Or, as Kant would have put it, what motivates us to be rational is not simply the obligation to believe the (apparent) truth, but also the obligation to form a set of beliefs that constitute a coherent “experience” of the world in the full Kantian sense of this term.

It is this obligation that explains why we should in fact care about being epistemically rational in the sense under discussion. And I think that this is an obligation that we all endorse from a considered, epistemic point of view. This means that, even from an epistemic point of view, our fundamental concerns are not exhausted by a concern for the truth - rather the basic obligations that motivate us in trying to be epistemically rational involve matters like coherence and explanatory power - matters that have obvious value from a epistemic point of view, even though they have no necessary connection with the truth.

To my mind, this is the one of the most important lessons of a philosophical encounter with skepticism. Part of what the skeptic’s scenarios teach us, if we trust our basic intuitions about them, is that there is no general, necessary connection between what we intuitively take to be epistemically rational and what is reliable. The skeptic’s response to this discovery is to retreat to a highly cautious conception of epistemic rationality that secures a connection of this sort, but only by accepting that it is epistemically irrational to believe P if we are not certain that P. The externalist, on the other, responds to these cases by rejecting the internalist intuition that the rational status of our beliefs supervenes on facts that we have internal access to. By doing so, the externalist also secures a necessary connection between rationality and reliability, but only at the cost of our internalist intuitions. Far better than either of these options, I believe, is the option of retaining our internalist intuitions, while accepting that there is more to epistemic rationality than reliability alone. Of course, this does involve an important concession to the skeptic - namely, that the value of epistemic rationality involves obligations other than those connected with the truth. But it seems to me that this involves no great alteration to our intuitive understanding of epistemic rationality. For our ordinary conception of epistemic rationality was

never limited to the obligation to believe the truth *on its own*.

Appendix: Obligations under Conditions of Uncertainty and Risk

In this essay, I've been concerned with arguments of the following general form:

- (1) X'ing is required in order for us to A.
- (2) A'ing is indefeasibly rationally obligatory.
- (3) If A'ing is rationally obligatory, and X'ing is required for us to A, then X'ing in order to A is also rationally obligatory.
- (4) Therefore, X'ing in order to A is rationally obligatory.

Here I want to consider a line of objection to this sort of argument that focuses on the question of what our obligations are under conditions of uncertainty.⁴⁹ To see the sort of objection I have in mind, consider the following case:

(BUTTON) Suppose you are a political prisoner of a despotic regime. Normal methods of torture having failed, your interrogators decide to break you by putting you through a grueling series of moral dilemmas and decision problems, the first of which is the following. You are placed before a button and are told that if you do nothing, one of your fellow political prisoners will be tortured to death. But if you do push this button, the following will be true. First, a second fellow prisoner will be tortured for five minutes, with no lasting damage. And, second, there will be a certain (unspecified) chance that the first fellow prisoner will be spared. Suppose that you have come to trust your interrogators completely. Should you push the button? Are you obligated to do so?

In this case, we are asked to consider whether one is obligated to perform an action (pushing the button), which is the necessary means towards accomplishing something that one is obligated to achieve if possible (saving the life of a fellow prisoner) but which will also have negative effects that are quite serious. Now if one knew that pushing the button would save the life of your fellow prisoner, then doing so would be obligatory. (So the negative effects in question *on their own* are not sufficient to defeat one's obligation.) But given that the probability that this will be the case is completely unknown to one, is one really obligated (or even permitted) to push it?

Some may have the intuition that one is *not* obligated to push the button under the

⁴⁹Thanks to David Velleman for raising issues of this sort to me.

circumstances described. And if this is correct - that is, if one is not obligated to push the button, this case *would* amount to a counterexample to the argument schema above. For it would describe a case in which we have an (undefeated) obligation to A and in which X'ing is to only possible means to A, but we are not thereby obligated to X.

But it seems to me that the natural response to this case is to say that *even* under these conditions one is obligated to push the button. After all, imagine being challenged by the second prisoner who was tortured as to why you pushed the button. Doesn't the answer, "I pushed it because it was the only way to have a chance of saving the first prisoner's life," seem satisfactory? And then imagine being challenged by the ghost of the dead prisoner in the opposite way. Would the answer, "I didn't push the button because doing so was certain to cause someone pain," really be satisfactory? The more I reflect on this case, the more inclined I am to say that we are, in fact, obligated to push the button in it.

Still, I'm not sure that everyone will share my intuitions. But whatever we think of BUTTON, there is an important disanalogy between it and the cases I discuss above. For while these cases may be thought of as involving a sort of decision problem under conditions of uncertainty, they involve cases in which we are equally uncertain about the probability that X'ing will have positive effects (e.g. true beliefs) and the probability that X'ing will have negative ones (e.g. false ones). Thus, unlike in BUTTON, where we are uncertain only about some of the probabilities in question, in the cases of interest, all of the relevant probabilities are completely unknown to us. And this means that the role of this uncertainty will actually be much less significant than in BUTTON. After all, while the *certainty* that X'ing will have negative consequences *might* defeat our obligation to X when X is the necessary means towards an obligatory end but the connection between X and this end is uncertain, it is much less plausible that this is the case where the connection between X and these negative consequences is *just as uncertain* as the connection between X and the necessary end.

For example, consider:

(BUTTON2) Suppose you are placed before a button as above and are told that if you do nothing, one of your fellow political prisoners will be tortured to death. But if you do push this button, the following will be true. First, there will be a certain (unspecified) chance that a second fellow prisoner will be tortured for five minutes, with no lasting damage. And, second, there will be a certain (again unspecified) chance that the first fellow prisoner will be spared. Suppose that you have come to trust your interrogators

completely, should you push the button? Are you obligated to do so?

Here, the answer seems obvious: you are obligated to push the button.

In fact, the most worrisome case in this area seems to me to be one involving decision under conditions, not of uncertainty, but of risk (using these terms in their technical sense). For example, consider:

(BUTTON3) Suppose you are placed before a button as above and are told that if you do nothing, one of your fellow political prisoners will be tortured to death. But if you do push this button, the following will be true. First, a second fellow prisoner will be tortured for five minutes, with no lasting damage. And, second, there will be a 1/100000 chance (determined by some random process) that the first fellow prisoner will be spared. Suppose that you have come to trust your interrogators completely, should you push the button? Are you obligated to do so?

Here I think it is very natural to think that one should *not* press the button. But this case (unlike BUTTON2) is very different from the sorts of cases I am interested in. For BUTTON3 is a case in which we have very precise knowledge about all the relevant probabilities. So its relevance to the arguments considered above is slight. Still, it might be thought to be a counterexample to the argument scheme I am working with as stated above. And so it might be thought to require some revisions to this argument scheme, even if it does not affect the *use* of the argument scheme I am interested in.

In fact, I don't think even this is the case. For it seems plausible that the best description of our obligations in BUTTON3 is the following: First, we have a defeasible obligation to save our fellow prisoner's life. But this obligation is defeated by the combination of *two* factors: (i) the fact that trying to save the person's life will also cause another person very significant pain and (ii) the fact that even if we try to save the person's life, we are *very* unlikely to succeed.⁵⁰ So BUTTON3 does not mean the conditions laid out in premise 2 of the argument scheme - since in BUTTON3 we do not have an indefeasible obligation to save our fellow prisoner's life in the relevant sense. And so even BUTTON3 does not represent a counterexample to the argument I have been using.⁵¹

⁵⁰Here I am invoking something like a probabilistic version of ought implies can.

⁵¹This response won't, I think, work for the earlier two cases. For while it is plausible that our obligation to A may be defeated by the fact that trying to A would both have negative side-effects and be very unlikely to succeed, it seems much less plausible to me that this obligation would be defeated by the fact that trying to A would have negative side-effects and that we are uncertain how likely trying to A would be to succeed.