

CAN RELIGIOUS UNBELIEF BE PROPER FUNCTION RATIONAL?

Michael Czapkay Sudduth

I. Introduction

In *Warrant and Proper Function* Alvin Plantinga argued that a belief has warrant, roughly, just if it is produced by cognitive faculties functioning properly in a congenial environment according to a design plan successfully aimed at truth. In chapter 6 of *Warranted Christian Belief* (hereafter WCB) Plantinga draws on Thomas Aquinas and John Calvin to present a model of how theistic belief can have warrant in a basic way (i.e., without propositional evidence). Plantinga seems to affirm the following proposition, a more recent variant of his earlier proper basicity thesis:

[P1] There are circumstances C such that, given any human person S, if S is in C and S's (relevant) truth-aimed cognitive faculties are functioning properly, then S holds a firm basic theistic belief.¹

By "circumstances" I will understand any of the many widely realized experiential conditions that Plantinga points out in chapter 6 of WCB, things like the starry night sky, the crashing waves of the ocean, the majestic grandeur of the Vermont Mountains, and the lovely melody of a Bach Concerto. Plantinga maintains that these sorts of circumstances trigger the formation of various kinds of theistic beliefs: God is present, is powerful, is forgiving, etc. (all of which self-evidently entail God exists). By "relevant" cognitive faculties I refer to Calvin's *sensus divinitatis*, which Plantinga takes to be a natural faculty or mechanism that is responsible for producing various theistic beliefs. More specifically, "the *sensus divinitatis* is a disposition or set of dispositions to form theistic belief in various circumstances, in response to the sorts of conditions or stimuli that trigger the working of this sense of divinity" (*WCB_M*, chapter 6).² The experiential circumstances, of course, are not evidences that are taken as premises from which theistic beliefs are derived by a process of inference. The circumstances are simply occasions that trigger the formation of theistic belief, so theistic belief is basic (i.e., not held on the evidential basis of other beliefs). Plantinga further thinks that when the *sensus divinitatis* is functioning properly we will hold a *firm* theistic belief. More precisely, if a person's (relevant) cognitive faculties are functioning properly and she is in any of the widely realized experiential circumstances (specified by the design plan), she will firmly hold some theistic belief.

Plantinga also claims in chapter 6 that if theism is true then [P1] or something close to it is likely to be a true epistemological proposition. (He would also add that there *are* in fact some people who find themselves in the relevant circumstances and whose relevant cognitive faculties are functioning properly, and thus who hold a firm basic belief in God). Plantinga presents the argument as follows. If one thinks of humans as created by God, one will not think of theistic belief as the product of belief forming processes aimed at something other than truth, much less as the product of some intellectual defect. If God exists and has created human persons in His image for the purpose of entering into communion with Him, then it is natural to think that God desires us to hold true beliefs about Him and our duties to Him. So He would probably create us in such a way that we can achieve this cognitive goal. This seems likely if theism is true. But then it is also likely that the faculties that produce theistic belief have been aimed at this end by their designer. In that case, theistic belief is the product of cognitive faculties functioning properly according to a design plan successfully aimed at truth.

Plantinga's notion of proper function rationality (hereafter PF-rationality) raises an interesting question: Can religious *unbelief* be PF-rational? (By "religious unbelief" I will understand, unless otherwise noted, either withholding theistic belief or believing the negation of theism). Plantinga thinks that if we do not assume a theistic metaphysics, an affirmative answer could be given to this question. If Plantinga is right about this a more interesting question would be whether religious unbelief could be PF-rational if theism is true. Plantinga argues that it cannot, at least given his model of proper basicity. "Unbelief," he says, "is a result of dysfunction, or brokenness, failure to function properly, or impedance of rational faculties" (*WCB_{M*}*, chapter 6). Plantinga presents some very interesting arguments for this claim in chapter 14 of *WCB*. However, I think the arguments are deeply problematic and entail significant difficulties for some of Plantinga's other important claims in religious epistemology. To see this I begin by developing a case for the PF-rationality of religious unbelief given some of Plantinga's own epistemological principles.

II. A Case for the Proper Function Rationality of Religious Unbelief

A necessary condition of warrant is the proper functioning of one's cognitive equipment. Fundamental to proper function is the cognitive design plan - a set of blue prints or specifications for a well-formed, properly functioning human cognitive system. Since the kind of specifications relevant for warrant are truth-oriented, they are specifications for that segment of the cognitive design plan that has as its purpose the production of true beliefs (as opposed to non-alethic purposes, such as survival or relief from suffering). The design plan specifies what the appropriate doxastic response of our cognitive faculties should be in a wide range of circumstances to achieve this alethic goal in a reliable manner.³

The design plan for humans also includes what Plantinga calls a *defeater system*, a cognitive subsystem that is designed to regulate modifications in a person's noetic structure given new experiences and the acquisition of new beliefs which come with social exposure, mental maturation, and education. In short, these are specifications as to the correct or proper ways of changing beliefs in response to experience (doxastic and otherwise). Generally, if a person acquires a defeater for a belief B, then there ought to be a certain kind of revision in the person's noetic structure. In some cases the person ought not to hold B with the same degree of firmness (partial defeat). In other cases the proper response is not to hold B at all (complete defeat). Defeaters, partial or complete, may be reasons for no longer holding B (i.e., undercutters) or reasons for holding a belief incompatible with B (i.e., rebutters). (Although Plantinga has elsewhere taken defeaters to include non-doxastic experiences, I will confine myself in this paper to defeaters construed as other beliefs of a person). A belief D is a defeater for a belief B of some person S just if, given S's noetic structure, S cannot rationally hold B (at least to some degree) given that S also holds D.⁴

Plantinga cites a couple of examples of defeaters. I see in the distance what appears to be a sheep in the field and form the belief that there is a sheep in the field. The next day someone comes along whom I know is the owner of the field and he tells me that there are no sheep in the field. I have acquired, via testimony, a *rebutting* defeater for the belief that there was a sheep in the field because I have acquired a reason for supposing that there was no sheep in the field. But there are also *undercutting* defeaters. A person enters a factory and sees an assembly line on which there are a number of widgets. These appear red. Being appeared to red-widgetly, the person forms the belief that there are red widgets on the assembly line. The shop superintendent then informs the person that the widgets are being irradiated by an infra red light, thereby enabling the detection of otherwise undetectable hairline cracks. Here the person merely loses his reason for supposing that the widgets are red. Both of these cases presents us with a person (i) who holds some belief B at time t_1 , (ii) whose noetic structure undergoes modification with respect to B at time t_2 after it comes to include the additional belief D, and (iii) whose relevant cognitive faculties (we correctly judge) are functioning properly at t_1 through t_2 .⁵

Since the defeater-system is aimed at the production and sustenance of true beliefs, it forms an essential thread in the array of our rational cognitive faculties. The proper function requirement for warrant, then, extends to the proper functioning one's defeater system, what we might call defeater-system PF-rationality. A person S's belief B is warranted only if S is PF-rational in holding B (to the degree that S does). And S is PF-rational in holding B (to the degree that S does) only if the relevant portions of S's defeater system are functioning properly and free from impedance, functioning such that the appropriate revisions take place with the acquisition of any defeaters against B.⁶ If there was no noetic modification in the above cases, then the person's cognitive state would not be PF-rational. Consequently, the belief B in each case would lack warrant (to some degree).

So Plantinga holds to what we can call the *rationality defeater principle*:

[DP] Given any person S and any belief B, if S acquires a(n) (undefeated) defeater D for B, then S is no longer rational in holding B (at least not with the same degree of firmness).⁷

It follows from [DP] that *if* a person acquired a(n) (undefeated) defeater for theistic belief, then holding theistic belief (or doing so firmly) would be PF-irrational and not holding theistic belief (or not doing so firmly) would be PF-rational.

Consider a case not too far removed from one that Plantinga himself introduces.⁸ Lisa has been raised in a Christian family. During her youth she holds her theistic belief in a basic way, but in her later teenage years her theistic belief isn't as strong. The cares of college life, sexual indulgence, and late-night parties slowly erode her thoughts of God. While a senior in college, she is exposed to Sigmund Freud's idea of wish fulfillment. She becomes convinced that the belief she had in an invisible friend called Merlin while a young girl was one such belief. Upon further reflection, though, she sees that her belief in God is significantly analogous to the belief she once had in the invisible Merlin. So she comes to believe that (p) her belief in God is really the product of wish fulfillment, a convenient defense mechanism against the hostile forces of one's environment. Her readings in Freud confirm this. Moreover, she also believes that (q) the objective probability of a belief being true given that it is produced by wish fulfillment is either low or inscrutable. (I say "or inscrutable" here because perhaps she is simply agnostic about the probability of a belief being true given that it was produced by wish fulfillment, rather than estimating that probability to be low). She then believes that the objective probability of her theistic belief being true is either low or inscrutable. Lisa has acquired an undercutting defeater for her theistic belief. If her defeater is partial and itself undefeated (as might be the case if she didn't hold either p or q very firmly), then the rational thing to do would be to hold her theistic belief less firmly than she did before acquiring this defeater. Perhaps her defeater is complete and itself undefeated (e.g., she has great enthusiasm for projective theories of religious belief or alternatively her theistic belief is very weak before encountering p and q), then if she is PF-rational she will no longer hold her theistic belief at all.⁹

Then there is Elvis' problem of evil. Elvis is a very devout believer in God raised in an Assembly of God Church in Tupelo, Mississippi. While in the Army, Elvis' mother Gladys dies of a heart attack. As a result, Elvis loses his faith. He reasons thus: (A) if there is a God, then God would not permit my mother to die unless he had a good reason for doing so, (B) my mother is dead and God had no reason for permitting her to die, therefore (C) there is no God. Elvis has acquired a rebutting defeater. If Elvis holds (A) and (B) very firmly and sees the deductive entailment (C), then Elvis probably has a complete rebutting defeater for his theistic belief. The rational thing to do is no longer hold theistic belief (maybe even hold the negation of theism). If he continued to hold his theistic belief, even in a less than firm manner, he would not be PF-rational. On the other hand, it is possible that Elvis does not hold the second conjunct of (B) very firmly and so in fact has simply acquired a partial defeater. The rational thing to do is no longer hold his theistic belief with the same degree of firmness. If Elvis continued to hold his theistic belief with the same degree of firmness, that belief would not be PF-rational. Since PF-rationality is necessary for warrant, Elvis' theistic belief would not be warranted. And since warrant is necessary for knowledge, Elvis would not know that God exists.

It is important to point out that the argument here for the PF-rationality of religious unbelief is logically consistent with [P1]. What follows from [P1] is that if a person is *in the relevant circumstances* and does not hold a firm theistic belief, holds no theistic belief, or believes the negation of theism, he suffers from some cognitive malfunction, or perhaps his cognitive state has been produced by something other than truth-aimed cognitive faculties. In addition to circumstances C (that call for firm theistic belief), there are plausibly circumstances C* which include having an undefeated defeater for theistic belief. Here, depending on the actual defeater, the appropriate doxastic response for a reasonable person will be withholding theistic belief, holding the negation of theism, or merely holding a less than firm theistic belief. The relevant analogy here can be taken from the conditions that govern the PF-rationality of sensory perceptual, testimonial, and memorial beliefs. If a circumstance includes being appeared to rain, then I am PF-rational in holding the belief that it is raining outside, unless of course the relevant circumstance includes my having defeaters for such beliefs. So even if [P1] is likely if theism is true, it would not follow that it is likely that religious unbelief is PF-irrational if theism is true. This point is significant and I will return to it in section IV.

III. Degrees of Rationality and Irrationality Transference

Although Plantinga allows for the possibility that Lisa and Elvis each acquires a defeater for theistic belief, he maintains that the cognitive state of religious unbelief is not therefore PF-rational.^{10,11} Plantinga claims that if a person has a defeater for theistic belief it is because there is malfunction elsewhere in the person's cognitive system. This is developed in chapter 14 of *WCB*. Plantinga's main contention here is that if a person acquired a defeater for theistic belief, then there would have to be cognitive malfunction elsewhere in the person's cognitive system, most likely in the *sensus divinitatis* itself. With reference to the problem of evil as a defeater for theistic belief, Plantinga writes: "What it is important to see, here, is that if she has a defeater, it is only because of a failure of rationality somewhere in her noetic structure (perhaps there is dysfunction with respect to the *sensus divinitatis*)." (*WCB_{M*}*, ch. 14).

I think we can provide an initial formulation of Plantinga's argument here by laying down what appears to be the argument's two main premises:

[P2] Given any fallen person S(f), if S(f) acquires a defeater D for some theistic belief T, then S(f)'s acquiring D depends on cognitive malfunction in another relevant cognitive module.

[P3] Given any fallen person S(f), if S(f)'s acquiring a defeater D for some theistic belief T depends on cognitive malfunction in another relevant cognitive module, then the cognitive state of religious unbelief is not PF-rational.

Let's begin by addressing the clause "S(f)'s acquiring D depends on cognitive malfunction in another relevant cognitive module." One way of unpacking these premises is to parse this clause in terms of "irrational input to the defeater system." Plantinga does not explicitly state this, but I think it is a plausible way of thinking of the notion of a defeater's depending on cognitive malfunction. In other words, defeating reasons may include a belief that was produced by a malfunctioning cognitive faculty. If we go this route, Plantinga's argument would involve something like an irrationality transference principle, according to which irrationality is transferred from an irrational defeating reason to the cognitive state based on it. If belief B is based on belief A, and A is PF-irrational, then B is PF-irrational. The PF-rationality of a cognitive state based on defeaters would then require that defeating reasons not be produced by any defective faculties. Religious unbelief then would be PF-irrational because defeaters for it in some way take as input beliefs that have been produced by some dysfunctional faculty.

Two important distinctions need to be made here. First, we need to distinguish between the cognitive state of religious unbelief and the defeater that is in some way responsible for that cognitive state. We can evaluate each in terms of PF-rationality. Secondly, we need to distinguish different degrees of PF-rationality based on something like *depth of rationality ingression*. In a foundationalist scheme beliefs have different degrees of depth of ingression depending on where they are located within a person's noetic structure (specifically, how close they are to the foundations and how much of the doxastic content of the superstructure depends on them). Similarly, we can say that there is a depth of ingression with respect to the PF-rationality of beliefs, and that this is a determinant, maybe the primary determinant, of the degree of PF-rationality of a cognitive state. Suppose that a malfunction in some cognitive faculty f1 produces some belief p that, together with a belief q produced by a properly functioning faculty f2, functions as input to the defeater system. Suppose further that, given these conditions, if the defeater system is functioning properly it would produce the belief r. The suggestion is that the cognitive state <belief that r> is PF-rational *to some degree*, and its degree of PF-rationality is greater than the degree of PF-rationality for the cognitive state <belief that p>. The latter is directly produced by cognitive malfunction and has what we might call a zeroeth degree of PF-rationality ingression. The former is the product of some properly functioning faculties (two to be exact), though it also includes an irrational belief as input to the defeater system. It has a degree of PF-rationality ingression greater than zero, but not maximal. A maximally PF-rational cognitive state would be one that is produced *solely* by truth aimed, properly functioning cognitive faculties and – if nonbasic – is located on an inferential path which has no irrational beliefs.

It would seem, then, that the irrationality transference principle is mistaken. Religious unbelief might have a degree of PF-rationality ingression greater than zero, even if the defeating reasons producing it do not. We can plausibly read Elvis and Lisa's cognitive states of religious unbelief at least as instances of less than maximal PF-rationality. In fact, one might argue that their defeaters are maximally rational.¹² But the important point is that even if the defeaters were less than maximally rational, they would still be epistemically significant. Plantinga himself seems to concur, for he admits in his unpublished "Naturalism Defeated" and in *WCB* (chapter 11) that an irrational belief can function as a rationality defeater. Suppose I believe that all cats are benign creatures, but then due to some rare feline phobia I come to believe that my cat Salem crawls to my face at night with the intention of suffocating me (and that perhaps he conspires in this with other neighborhood cats). According to Plantinga, I thereby acquire a rationality defeater for my previous belief that all cats are kind and benign creatures, even though the defeating reasons have a zeroeth degree of PF-rationality ingression, with the malfunction occurring in another relevant cognitive module. As a result of acquiring this defeater, not only is my previous optimistic feline belief no longer rational but that belief no longer has warrant (at least not to the same degree).

This suggests that a particular doxastic state can be epistemically significant and PF-rational even if that doxastic state supervenes on cognitive malfunction by way of irrational input to the defeater system. In the above case, is my doxastic state maximally rational? No. The input to the defeater system includes beliefs that were produced by cognitive malfunction. In fact, the belief has little if anything in the way of warrant.

Something is wrong with me, but if I continued to hold my optimistic feline belief *two things* would be wrong with me, as my failure to modify my optimistic feline belief indicates a second cognitive defect (related to the proper functioning of my defeater system). So it seems that the human design plan at least *sometimes* specifies that withholding belief (or holding a belief less firmly) is called for even when this depends on dysfunction somewhere else in one's cognitive system. Since a belief can lose warrant if a person acquires a defeater for it, an irrational belief can defeat warrant as well. So nothing in principle robs a cognitive state of epistemic significance just because it was produced by cognitive processes that include a less than maximally rational defeater. Of course, it does not necessarily follow that the same thing is true for theistic belief. But on what grounds can we affirm this of other beliefs and deny it for theistic belief in particular without the appearance of arbitrary epistemic partiality?

IV. Proper Basicity, Relevant Analogs, and Rationality Defeaters

Well, perhaps we have moved too quickly. I have been assuming that Plantinga's argument for the irrationality of religious unbelief depends on a claim about irrational *input* to the defeater system, where that irrationality is allegedly transferred to the cognitive state of religious unbelief. But parsing [P2] and [P3] in this way is not the only way to go. We might simply understand a defeater's depending on malfunction in another *relevant* cognitive module to refer to the malfunctioning of a cognitive faculty that is designed to produce theistic belief. On this way of looking at things, Plantinga could affirm the possibility of maximally rational defeaters for theistic belief but still deny that religious unbelief is PF-rational. Plantinga could respond that even if Elvis or Lisa each has a maximally rational defeater for theistic belief, the fact that they end up in a state of religious unbelief shows us that *something* is wrong. He claims that this points to a malfunctioning *sensus divinitatis*. Plantinga writes: "On the extended A/C model, the *sensus divinitatis* is among our cognitive faculties or processes; if it is functioning properly in S, then the belief that there is such a person as God will automatically have warrant" (*WCB_{M*}*, chapter 14). Again he writes of the *sensus divinitatis*: "someone in whom this process was functioning properly would have an intimate, detailed, vivid and explicit knowledge of God. . . . but the idea that perhaps there just wasn't any such person as God would no doubt not so much as cross her mind" (*WCB_{M*}*, chapter 14). So the irrationality of religious unbelief is not due to irrational input to the defeater system, but the malfunctioning of a faculty designed to produce theistic belief.

Plantinga then seems to be affirming the following proposition (as an entailment of [P2]):

[P4] Given any fallen human person S(f), S(f) has some cognitive faculty F (i.e., the *sensus divinitatis*), such that if F is functioning properly, S(f) holds a firm theistic belief T.

It follows from [P4] that Elvis's holding a less than firm belief that T, or failing to hold T altogether, is an indication of cognitive malfunction (or impedance), *even if Elvis has a defeater for T*. Elvis acquires a defeater for theistic belief, at least in part, because the *sensus divinitatis* (hereafter SD) is malfunctioning, not because another faculty produces an irrational belief that forms part of the defeater. If the SD were functioning properly, then a person would probably not acquire a defeater for theistic belief in the first place (and if he did, he would have a defeater-defeater for this defeater in the form of the output of the SD). On the other hand, if the SD is damaged or non-operational, a person might not have a firm belief in God, thereby rendering him susceptible to acquiring a defeater for theistic belief and moving further into religious unbelief.

Plantinga has offered two analogies to support this position.¹³ First, suppose that a student acquires a good argument for believing that he no longer exists, say under the influence of his persuasive philosophy professor. Would giving up his belief in his own existence be rational? No, Plantinga says; not all. Secondly, suppose a woman comes to believe that her cognitive faculties are unreliable because she thinks she has contracted mad-cow disease. According to Plantinga she has a rationality defeater for all her beliefs

since she has a reason for supposing that her cognitive faculties are not reliable. Perhaps the defeater is even a maximally rational one (i.e., does not depend on irrational input to the defeater system). If we traced its path through the woman's noetic structure we would not find a defective segment anywhere in route to the foundations. But Plantinga affirms that proper function would *not* require that she give up all her beliefs. "That way lies sheer madness." I take Plantinga's point here to be that the design plan does not stipulate universal doxastic abstention under any circumstances. So no matter what reasons a person acquired for thinking that her faculties were not reliable, she ought not to give up all her beliefs. (Of course, since Plantinga construes "having a defeater in some circumstance" as determined by the design plan, if the design plan doesn't make provision for withholding all beliefs under any circumstance, it would seem that the woman can't even get a defeater for all her beliefs). Plantinga would have us regard a cognitive state of religious unbelief in the presence of defeaters as analogous to universal doxastic abstention in the presence of a defeater for the reliability of our cognitive faculties. If the latter is not a PF-rational cognitive state, neither is the former. As universal doxastic abstention, or no longer believing in one's own existence, is not PF-rational given the acquisition of any set of defeating reasons, so also religious unbelief is not PF-rational given the acquisition of any set of defeating reasons for theistic belief.

The argument against the PF-rationality of religious unbelief comes down to the plausibility of [P4], and this question in turn depends on the extent to which [P4] can draw support from the kind of analogies that Plantinga presents. As I see it, there are problems here. The most serious problem is that [P4], and the supporting analogical argument, entails a significantly different formulation of the proper basicity thesis than the one Plantinga endorses elsewhere. For instance, in chapter 6 of *WCB*, when Plantinga lays out the model for warranted basic theistic belief, the proper functioning of the SD is always associated with or qualified by a definite range of experiential input, in a way analogous to the formation of sensory perceptual beliefs.

Plantinga writes:

There is a kind of faculty or cognitive mechanism. . . which in a wide variety of circumstances produces beliefs about God. . . . Under these circumstances. . . these beliefs are formed in us. . . . The *sensus divinitatis* is a disposition or set of dispositions to form theistic belief in various circumstances, in response to the sorts of conditions or stimuli that trigger the workings of this sense of divinity. . . . According to the model, therefore, there are many circumstances, and circumstances of many different kinds, that call forth or occasion theistic belief. Here the *sensus divinitatis* resembles other belief-producing faculties or mechanisms. If we wish to think in terms of the overworked functional analogy, we can think of the *sensus divinitatis* too as an input-output device: it takes the circumstances mentioned above as input and issues output theistic beliefs, beliefs about God" (*WCB*_{M*}, chapter 6)

But the formulation that emerges in the discussion on defeaters omits the circumstance relative nature of the theistic belief-forming process and its dependence on certain experiential input that triggers theistic belief. This is no minor variation. The earlier view, formulated in [P1], is logically consistent with a person's SD functioning properly and the person not holding theistic belief (or not holding it firmly), for in [P1] the proper functioning of the SD is contextually situated in a limited range of circumstances. Plantinga's argument for the dependence of defeaters on SD malfunction and the irrationality of religious unbelief seems to depend in a most crucial way on [P4]. But it follows from [P4] that religious unbelief is PF-irrational and theistic belief PF-rational in *any* circumstance, not just those referred to in [P1]. Although Plantinga is certainly free to develop his own epistemological model for warranted theistic belief, the problem is that in fact he has two different models, [P1] and [P4]. The first is consistent with the PF-rationality of religious unbelief in some circumstances; the second is not.

Can't Plantinga just drop [P1] and assert [P4]? Well, he certainly could, but should he? The likelihood of [P1] given the truth of theism is itself controversial, as it is not clear how the truth of theism makes it likely that we should hold theistic belief in a basic (as opposed to nonbasic) way. [P4] would be even more controversial. It would require affirming that God not only wants us to have true beliefs about him that are

formed in a basic way, but that it is likely that he wants us to have basic beliefs about him (or at least his existence) that could not be subject to defeat, except on the condition of irrationality.

More importantly, Plantinga's taking [P4] as the model will really depend on the extent to which Plantinga wants to draw crucial analogies between theistic belief and certain other paradigmatic properly basic beliefs. For instance, when it comes to defending basic theistic belief against the charge of fideism (in chapter 10 of *WCB*), Plantinga is quick to establish the similarity between basic theistic belief and other basic beliefs that are not immune to defeat just because they are basic. Plantinga says: "Theistic belief would certainly not be immune to argument and defeat just by virtue of being basic. In this, theistic belief only resembles other kinds of beliefs accepted in the basic way" (*WCB*_{M*}, chapter 10). His three examples are beliefs accepted on testimony, sensory perceptual beliefs, and Gottlob Frege's belief that for every property or condition, there exists the set of just those things. Plantinga seems to be saying that the design plan for basic theistic belief is similar to the design plan for some other basic beliefs in that they share the property of being susceptible to defeat, and defeat without irrationality anywhere else in one's cognitive system.¹⁴

This tells us something important about the faculty responsible for theistic belief, or at least how Plantinga is thinking of it. In the case of sensory perceptual beliefs, the relevant cognitive module responsible for the formation of such beliefs is designed to function properly and yield certain beliefs as output *given certain experiential input* and given the exclusion of sufficient reasons to the contrary. Upon being appeared to rainily, I form the belief that it is raining outside. If the relevant portions of my cognitive system are functioning properly, then I will hold the belief that it is raining outside. But note. My failure to believe it is raining outside is not by itself an indication of cognitive malfunction. I might not be in the relevant circumstance that is specified by the design plan (i.e., being appeared to rainily). Alternatively, perhaps the circumstance I am in includes being appeared to rainily as well as reasons for supposing either that this belief is false or that its ground is inadequate. Take the situation where I hear "pitter-patter, trickle, drip, drip" coming from outside and I see drops of water hitting against my kitchen window. I form the belief it is raining outside. But then a friend of mine comes in (perhaps drenched with water) and tells me that there's some crazy old man outside who's spraying water everywhere (and I have no reason to believe that my friend is lying to me). So I no longer believe that it is raining outside. Is there any cognitive malfunction here? No. My holding a less than firm belief that it is raining outside is compatible with the proper functioning of the cognitive faculties which produced that belief in me in the first place. In fact, I would be PF-rational in withholding the belief even if it *was* raining outside (perhaps the old man always sprays water on people when it is raining outside). And the same argument could be made for testimonial beliefs and the Frege example.¹⁵

Now Plantinga is surely correct about one thing. It does seem reasonable to suppose that the design plan rules out some cognitive states as PF-rational, even if the state is grounded, allegedly at any rate, in a defeater.

Plantinga's two examples are appropriate candidates here. But we can dig deeper. One of the modules of our cognitive establishment is responsible for what we might broadly call self-knowledge. It produces beliefs like "I exist" and a broad range of introspective beliefs about one's current states of consciousness (e.g., I am in pain, I am being appeared to redly). It is widely held, though not beyond controversy, that such beliefs possess certain epistemic immunities: immunity from doubt (indubitability), error (infallibility), or revision (incorrigibility). The concept of defeat only seems appropriate when beliefs do not have such properties. The notion that such beliefs involve some kind of privileged epistemic access makes it difficult to see how a person could acquire a defeater for them and actually come to doubt them without us supposing that something has gone terribly wrong with the person. In the case of my believing that I'm being appeared to redly, the self-presenting doxastic state is produced from the input <being appeared to redly>. If the faculty (or sub-faculties) responsible for introspective beliefs is functioning properly then whenever we have the corresponding experiential input, then we hold the belief that we are being appeared to a certain way. Unlike the case of a sensory perceptual belief (e.g., it is raining outside), there are no circumstances that could *include* the relevant experiential input to the belief forming mechanism without yielding the relevant introspective belief, at least not without malfunction somewhere.

Even if a person acquired apparent defeating reasons for the introspective belief, she would (if rational) continue to hold this belief with the same degree of firmness. In the case of belief in my own existence, it is reasonable to suppose that if my cognitive faculties are functioning properly then I will never withhold the belief or hold its negation.

This can also help us think about Plantinga's woman who doubts the reliability of her cognitive faculties because she thinks she has contracted mad cow disease. What exactly happens here? The woman acquires reasons for supposing that her cognitive faculties are unreliable. But the woman must assume that some of her cognitive faculties *are* reliable if she has any reason for supposing that her cognitive faculties are not reliable. Well, perhaps the woman does not consciously assume this, but it certainly seems presupposed in a significant sense by her having a defeater in this situation and recognizing it. Without at least an implicit commitment to or belief in the reliability of at least some of her cognitive processes, specifically the ones responsible for producing her alleged defeating reasons, she really has no good reason for supposing that her cognitive faculties are unreliable. We could then handle this counterexample by saying that the design plan for our cognitive establishment precludes as instances of proper function all cognitive processes that involve this kind of self-referential incoherence or self-defeat, as it runs contrary to the truth goal of the cognitive design plan. Perhaps we can generalize a bit and say that if the holding of any apparently defeating reasons D for some belief B entails the holding of the defeatee, then D cannot be a defeater for B. Suppose, though, that a person believed that <if God does not exist, then one's cognitive faculties are not reliable>. If one acquired a rebutting defeater for God's existence, then one would probably be in a similar circumstance as the woman above. Here the holding of one's defeating reasons would be inconsistent with withholding the defeatee. But this is obviously a special case and would not be true of just any defeater for theistic belief.

It is reasonable to suppose that some of the faculties of our noetic establishment are such that if we are in some circumstance (such as being appeared to a certain way), then nothing could be added to the circumstance (in the way of reasons to the contrary) to make withholding the corresponding belief PF-rational. Similarly, there are faculties that, if functioning properly, preclude the withholding of certain belief(s) under any circumstances. We might suppose that in both of these cases the original belief is actually a defeater against any potential defeater. If a philosophy professor gives me a good argument for the proposition that I do not exist, it would seem that my belief that I do exist (held firmly and PF-rational) has so much more by way of warrant that it could simply crush such an argument. I have no reason to take the argument seriously at all. Also, there will be certain alleged defeating reasons for some beliefs (like belief in the reliability of our cognitive faculties) that implicitly involve a commitment to the truth of the defeatee. It would not be rational to withhold belief in such cases.

Operating inductively from these cases I think we can frame criteria for the kinds of beliefs that cannot be withheld without supposing that the faculty that produces them is malfunctioning. Since Plantinga has elsewhere suggested such an approach to understand what beliefs are properly basic, it would be reasonable to take such an approach in trying to determine what doxastic withholdings can be PF-rational and which not. First, these would be beliefs that have epistemic immunities that render the notion of defeat logically inapplicable to them. Secondly, there will be doxastic states that are self-referentially incoherent or where the grounds for withholding belief presupposes a person's commitment to the truth of the defeatee. Perhaps further examples could be drawn from the cognitive modules responsible for our belief in some *a priori* propositions (e.g., analytically true propositions) and all doxastic withholdings in the case inconsistent propositions (that we see to be inconsistent at any rate). But theistic belief really doesn't satisfy these criteria. So a person's acquiring defeaters for theistic belief can be PF-rational since theistic belief does not share the necessary properties that would make it immune to malfunction-free defeat.

V. Conclusion

Warranted Christian Belief presents some strong claims about the irrationality of religious unbelief, based on the truth of theism and Plantinga's model of properly basic theistic belief. Plantinga claims that if theism

is true, then his proper basicity model (or something close to it) is likely to be true. I have argued that in fact Plantinga has two models, and that the PF-rationality of religious unbelief is logically inconsistent with only one of them, the model that includes [P2] and [P4]. But neither of these propositions appears likely if theism is true. In fact, these formulations seem most *unlikely* if other aspects of Plantinga's epistemology are correct or logically consistent. If these formulations were true, Plantinga would lose what has hitherto been an important component of his religious epistemology – the parity between basic theistic belief and other basic beliefs that are susceptible to defeat and for which there can be PF-rational doxastic withholdings. Denying this leaves Reformed epistemology quite vulnerable to the charge of fideism. Conversely, insulating Reformed epistemology from this charge would seem to require that there are circumstances in which religious unbelief can be PF-rational, even if theism is true. Theistic belief is not beyond rational appraisal or proper function rationality defeat.¹⁶

Saint Michael's College

NOTES

¹ This claim is significantly different from another formulation of the proper basicity thesis (also found in *Warranted Christian Belief*): theistic belief held in a basic way is proper function rational or warranted if theism is true. Here theistic belief and proper function are not situationally indexed or made relative to any set of circumstances. I will discuss the significance of this alternative formulation in section IV.

² References to *Warranted Christian Belief* (*WCB*) will be given parenthetically in the text by chapter. *WCB_{M*}* indicates that quotations are from Plantinga's 1997-98 unpublished manuscripts of *WCB*. Unfortunately, since this paper was completed in advance of the proofs for *WCB*, I am unable to supply page numbers.

³ The position is *externalist* since the proper functioning of one's cognitive system is not the sort of thing to which a person has introspective access or could come to know just by reflecting on one's mental states. Viewed this way warrant entails a kind of *externalist rationality*. In Plantinga's scheme, there are three closely related senses in which a belief can be externally rational. A belief is externally rational if it is the product of properly functioning cognitive faculties, that is, if there is an absence of cognitive dysfunction, impairment, disorder, or pathology in some person's holding B. But a particular doxastic state might be the product not of malfunction but of properly functioning cognitive faculties in some way overridden by certain emotional states (e.g., anger, ambition, lust). So a belief can also be externally rational in the sense that it was produced by properly functioning cognitive powers not impeded, inhibited, or overridden by emotions of a certain type. But external rationality is, more narrowly speaking, a matter of a belief being a deliverance of one's rational faculties, being produced by properly functioning (and unimpeded) cognitive faculties successfully aimed at truth. External rationality in this third sense is directly connected to warrant. In this paper I refer to it as *proper function (PF) rationality*. A belief will lack warrant if it is PF-irrational, and a belief is PF-irrational if it is produced by (i) cognitive malfunction, (ii) properly functioning cognitive faculties overridden by certain emotions, or (iii) properly functioning cognitive faculties not aimed at the production of true beliefs.

⁴ In chapter 11 of *WCB* Plantinga offers two definitions of a defeater. According to his preliminary definition: "(D) *D* is a defeater of *B* for *S* at *t* iff (1) *S*'s noetic structure *N* (i.e., *S*'s beliefs and experiences and salient relations among them) at *t* includes *B*, and *S* comes to believe *D* at *t*, and (2) any person (a) whose cognitive faculties are functioning properly in the relevant respects, (b) whose noetic structure is *N* and includes *B*, and (c) who comes at *t* to believe *D* but nothing else independent of or stronger than *D* would withhold *B* (or believe it less strongly)." However, Plantinga admits that this definition is faulty, for one might have a rationality defeater for some belief *B* while the proper functioning of non-truth aimed

cognitive faculties or processes requires the holding of B (e.g., for psychological comfort or survival). So Plantinga adds the notion of a "purely epistemic defeater" in contrast to a "defeater *simpliciter*": (D*) D is a purely epistemic defeater of B for S at t iff (1) S's noetic structure N at t includes B and S comes to believe D at t, and (2) any person S* (a) whose cognitive faculties are functioning properly in the relevant respects, (b) who is such that the bit of the design plan governing the sustaining of B in her noetic structure is successfully aimed at truth, (i.e., at the maximization of true belief and minimization of false belief) and nothing more, (c) whose noetic structure is N and includes B, and (d) who comes to believe D but nothing else independent of or stronger than D, would withhold B (or believe it less strongly)." In cases where the *only* processes governing the production or sustaining of a belief are truth aimed, a purely epistemic defeater will be a defeater *simpliciter*. I will assume this distinction throughout. For Plantinga's other discussions on defeaters, see "Naturalism Defeated" (December 1994 draft, unpublished) and "Reliabilism, Analyses, and Defeaters," *Philosophy and Phenomenological Research* 55 (1995), pp. 334-342.

⁵ In "Reliabilism, Analyses, and Defeaters," Plantinga says that "a defeater D for a proposition (for a person S) must be such that it lowers the epistemic probability of the prospective defeatee: it must be the case that the epistemic probability of the proposed defeatee on the conjunction of D with the relevant rest of S's noetic structure is lower than on that relevant rest alone" (p. 441). Of course, for Plantinga, the conditional epistemic probability of A on B is a matter of the degree to which a rational person will accept A given that she also accepts B, reflectively considers A in the light of B, and has no other source of warrant for A (or its denial). So whether a person has a defeater for some belief really depends on the specifications of the design plan. It isn't entirely clear how narrow (or broad) those specifications should be, and Plantinga formulates no general principles here but merely relies on examples that carry with them a good degree of intuitive plausibility. Presumably we could formulate general principles from an induction of such samples, much like the method Plantinga suggests for determining the criteria of proper basicity. I leave that project for another time.

⁶ The concept of defeat under consideration here is clearly externalist, since it concerns what can rationally be believed in the sense of cognitive proper function. So we can speak of such defeaters as *external rationality defeaters*. However, since the reasons that serve to defeat a belief include other *beliefs* of the person, they are the sort of thing to which a person has cognitive or introspective access upon reflection. The defeaters, then, are also internalist defeaters (perhaps what William Alston refers to under the rubric of perspectival internalism in his "An Internalist Externalism" in *Epistemic Justification* (Ithaca: Cornell University Press, 1989)). As such, they should be distinguished from defeaters construed as merely some external fact about the subject's environment or cognitive situation that negatively affects the positive epistemic status of a person's belief. There is also an important connection between defeaters, internalism, and warrant. A person's belief will fail to have warrant if it is produced by cognitive faculties that are not functioning properly, not aimed at truth, or if there is something awry in one's epistemic environment. Since these are all design plan specifications to which a person does not have introspective access, the kind of defeat they produce is externalist. What gets defeated is the *warrant* a belief has. But if I acquire an internalist rationality defeater for my belief that p and the design plan specifies that I hold my belief that p less firmly, but I continue to hold it with the same degree of firmness, my belief is not externally or PF-rational. So my belief will lack warrant for externalist reasons that involve the failure of my cognitive system to respond properly to certain internalist conditions, namely the acquisition of some new belief(s). Internalist *rationality* defeaters, then, (because they defeat what can rationally be believed in the sense of proper function) can also be externalist *warrant* defeaters (though not all warrant defeaters are rationality defeaters). This highlights a more general point about internalist rationality in relation to warrant. A belief is *internally rational* just if it is the appropriate doxastic response to what is given to a person by way of her previous beliefs and current experience. Warrant appears to require both internal and external rationality. For a discussion on the internalist and externalist contours of Plantinga's rationality defeaters, see my "The Internalist Character and Evidentialist Implications of Plantingian Defeaters," *The International Journal for Philosophy of Religion* (forthcoming 1999).

⁷ Defeaters often consist of more than one belief. Suppose I believe that (p) John cannot swim, and then I learn that (q) John is a lifeguard and I also believe (r) all lifeguards can swim. I have a rebutting defeater for my belief that p. But it isn't necessarily the case that my continuing to hold p along with q is irrational.

If I learn that (s) a special class of lifeguards need not be able to swim (e.g., perhaps they are spotters) and (t) John has a spotter insignia sticker on his car, then I could be PF-rational in continuing to hold p. Although I continue to hold one of the defeating reasons, the defeating force of the defeater (as a whole) has been neutralized by a reason for thinking that the other defeating reason (r) is false. This is an example of a defeater being defeated. So in [DP] we must say that the kind of defeater that defeats a person's rationally continuing to hold some belief is an *undefeated* defeater.

⁸ See *Warranted Christian Belief*, chapter 11; *Warrant and Proper Function* (New York: Oxford University Press, 1993), p. 229-231; and "Reliabilism, Analyses, and Defeaters," pp. 336-342.

⁹ Given the importance of this example, I offer two quotes from Plantinga's own version of the example. "Compare the case with a believer in God, who, perhaps through an injudicious reading of Freud, comes to think that religious unbelief generally and theistic belief in particular is almost always produced by wish fulfillment. Such beliefs, she now thinks, are not produced by cognitive faculties functioning properly in a congenial environment according to a design plan successfully aimed at truth; instead they are produced by wish fulfillment, which, while indeed it has a function, does not have the function of producing true beliefs. Suppose she considers the objective probability that wish fulfillment, as a belief producing mechanism, is reliable. She might quite properly estimate this probability as relatively low; alternatively she might also be equally agnostic about the probability that a belief should be true, given that it is produced by wish fulfillment. But then in either case she has a defeater for any belief she takes to be produced by the mechanism in question. . . . She. . . has an undercutting defeater for her belief in God; if that defeater remains itself undefeated and if she has no other source of evidence, then the rational course would be to reject belief in God. That is not say, of course, that she would in fact be *able* to do so; but it remains the rational course" (*Warrant and Proper Function*, pp. 230-231). In *WCB* Plantinga writes: "Projective theories like Freud's could be a defeater for theistic belief (and hence for Christianity) for *some* people. Suppose I believe very firmly that if theism is true, there couldn't be any coherent projective theories of religious or theistic belief; suppose I also accept theism, though not particularly firmly. Now suppose that I then come to believe that (F) Freud's theory (or some other projective theory) is indeed coherent. (F) will then be a defeater - perhaps a partial defeater - for my theistic belief; as long as I accept it and continue to accept the rest of my noetic structure (including the idea that theism is true only if there are no coherent projective theories of theistic belief), I can't rationally accept theism. Of course that idea is false; but a false belief can nonetheless serve as a defeater" (chapter 11).

¹⁰ I take Plantinga's position here to be a denial of the *external* rationality of religious unbelief, not a denial of its internal rationality. Although internal rationality seems to be necessary for PF-rationality and warrant, it is not sufficient. See endnote no. 6 for this distinction.

¹¹ Although Plantinga does concede that there can be defeaters for theistic belief (in a less than fully rational noetic structure), in chapter 14 he also argues that perhaps the problem of evil is **not** a defeater for theistic belief at all. Why? If the *sensus divinitatis* is damaged, then a person might very well come to believe that the existence of God is improbable given the facts of evil and fall into agnosticism or atheism. In this situation, though, one cognitive module (i.e., the module responsible for probability judgements) is functioning properly, but another module (i.e., the *sensus divinitatis*) is not. However, if the two modules are designed to function in tandem, how things go when in fact one is not functioning properly is not necessarily a part of the design plan. So religious unbelief is more properly viewed as an unintended by-product of malfunction in the SD plus proper function elsewhere, but since this need not be any part of the design plan, religious unbelief need not be construed as PF-rational.

¹² Lisa comes to believe that her belief in an imaginary friend while a young girl was really due to wish fulfillment. She also thinks that the objective probability of beliefs being true if they are produced by wish fulfillment is either low or inscrutable. She sees strong analogies between this belief and her theistic belief. She becomes convinced that her theistic belief is also the product of wish fulfillment and that its objective probability of being true is thus either low or inscrutable. Even if Lisa's premises here are false, it doesn't follow that Lisa's holding them is the result of cognitive malfunction. The crucial beliefs in Elvis's case are: (A) if there is a God, then God would not permit my mother to die unless he had a good reason for

doing so and (B) my mother is dead and God had no reason for permitting her to die. Perhaps the consequent of (A) is false, or maybe the second conjunct of (B) is false. But why suppose that Elvis' holding these beliefs is the result of cognitive malfunction. Maybe he has taken on the testimony of the preacher that (C) God has reasons for allowing evil and (D) if God has reasons for allowing evil then he either reveals them to us or otherwise consoles us by providing some outward sign. Suppose further that Elvis holds the true belief that God has provided neither signs nor reasons for allowing the death of Elvis' mother. Even if the preacher came to hold these beliefs because of some cognitive disorder, Elvis takes them on testimony and they form an essentially valid argument against the existence of God. And Elvis believes nothing else that could neutralize or otherwise defeat the defeater (even though clearly he could).

¹³ Plantinga mentioned the first analogy in correspondence and introduced the second analogy in response to an earlier draft of this paper presented at the Annual Meeting of the American Academy of Religion in San Francisco, California, November 22, 1997.

¹⁴ There's a closely related problem worth pursuing here that I'll only mention in passing. I'm not sure that Plantinga's definition of a defeater in chapter 11 of *WCB* is logically consistent with his claims in chapters 10, 11, and 14 that there *can* be defeaters for theistic belief. His definition (see endnote no. 4) requires for D to be a defeater for S's belief B that any person S* who holds B and comes to believe D (and has the same noetic structure) would withhold B if S*'s cognitive faculties are functioning *properly in the relevant respects*. Since Plantinga permits an irrational belief to serve as a defeater (even against a rational belief), he correctly notes that "relevant" proper function must be understood to permit cases where the defeater arises from cognitive malfunction and serves as irrational input to the defeater system. But Plantinga claims that if a person acquires a defeater for theistic belief it is only because the *sensus divinitatis* is not properly functioning, not necessarily because of irrational input to the defeater system originating from some other malfunctioning faculty. At the very least it is not clear that a malfunctioning *sensus divinitatis* is consistent with having a defeater for theistic belief. Does this not qualify as a relevant cognitive malfunction? A person S may acquire some other belief D (the putative defeater), but since any person S* whose cognitive faculties are functioning properly in the respects relevant to the truth-aimed production and sustenance of theistic belief would *not* withhold B (by virtue of the proper functioning of the *sensus divinitatis*), D is not a defeater against B for S. Thus, either there can be defeaters for theistic belief or there cannot. If there cannot be defeaters for theistic belief, Plantinga loses his rebutting argument against the charge of fideism. If there are defeaters for theistic belief, then (by Plantinga's definition) the design plan makes provision for withholding theistic belief in some circumstances. But Plantinga's account of the *sensus divinitatis* and his arguments for the irrationality of religious unbelief imply that the design plan makes no provision for the withholding of theistic belief. So it seems like there cannot be defeaters for theistic belief after all. Plantinga actually entertains the plausibility of this conclusion in chapter 14 of *WCB*. See endnote no. 11.

¹⁵ I should think that the same thing is true in the case of memory beliefs and belief in other minds. Memory beliefs are easy enough. I believe that my friends and I had grilled chicken on our camp-out last summer, but then I pull out the video and see that in fact we had hamburgers and hotdogs. But now think about belief in other minds. Suppose a man is frozen for hundreds of years through cryonics. During that time all human persons on earth are gradually replaced by look-alike/act-alike androids, a process perhaps necessitated by high levels of cosmic radiation that slowly destroyed the human race. When the ice-man thaws he initially holds his belief that there other minds. After all, he looks around and everyone looks and acts like humans and he has no reason to think otherwise, but he eventually uncovers evidence (in the form of videos and written documents spanning over a hundred years or so) which provides strong support for the proposition that all humans were eventually replaced by these android replicas. No longer believing in the existence of other minds, at least human ones, would seem to be PF-rational, even if the ice-man is fitted with a module that produces belief in other minds in *most* other circumstances. Neither that module, nor any other, seems to be malfunctioning in this instance. Moreover, it seems that the same sort of conclusion would follow if in fact his belief was false (perhaps there are a few last humans alive in caves somewhere in Arizona or the whole thing is some hi-tech, virtual reality experiment performed on the ice-man by the government).

¹⁶ An earlier version of this paper was presented at the Annual Meeting of the American Academy of Religion on November 22, 1997 in San Francisco, California. I am grateful to Alvin Plantinga for his comments on the earlier draft and the correspondence that generated the paper. Thanks also to Kelly Clark and Linda Zagzebski for helpful comments on the original draft of this paper.

Copyright © 1999