



Tópicos, Revista de Filosofía

ISSN: 0188-6649

kgonzale@up.edu.mx

Universidad Panamericana

México

Wilson Mulnix, Jennifer
Reliabilism and Demon World Victims
Tópicos, Revista de Filosofía, núm. 44, 2013, pp. 35-82
Universidad Panamericana
Distrito Federal, México

Available in: <http://www.redalyc.org/articulo.oa?id=323028517002>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

RELIABILISM AND DEMON WORLD VICTIMS

Jennifer Wilson Mulnix
University of Massachusetts, Dartmouth
jmulnix@umassd.edu

Abstract

This paper defends reliabilism against the classic demon world victim thought experiment. In doing so, I underscore two of its key alleged intuitions. I then articulate a host of varied responses open to the reliabilist, arguing that these readily available responses provide the reliabilist with a way to either accommodate or reject these initial intuitions about the demon world victim thought experiment, and in a way consistent with reliabilism. Thus, I conclude that the demon world thought experiment does not undercut reliability as the hallmark of epistemic justification.

Keywords: Epistemology, externalism, reliabilism, demon world victims.

Resumen

En este trabajo se defiende el fiabilismo frente al experimento mental clásico del “genio maligno”. Al hacerlo, enfatizo dos de sus supuestos clave; y después desarrollo una serie de variadas respuestas asequibles al fiabilista, mostrando que las mismas pueden ser útiles para explicar o negar las intuiciones iniciales de la propuesta del “genio maligno”, de una manera consistente con el fiabilismo. Mi conclusión es que el experimento del “genio maligno” no socava la fiabilidad como el sello distintivo de la justificación epistémica.

Palabras clave: Epistemología, externalismo, fiabilismo, “genio maligno”.

Recibido: 16-05-2012. Aceptado: 08-03-2013.

0. Introduction

This paper will defend reliabilism, one of the most prominent versions of externalism, against the classic demon world victim thought experiment.¹ In broad strokes, externalism maintains that a true belief is an instance of knowledge when the belief's content connects in the appropriate way with the part of the world that determines its truth, whether the subject who has the belief is aware of this connection or not.² If the appropriate connection is absent, then the given belief is not justified. But, according to many internalists, this violates our fundamental intuitions about epistemic rationality. For, how could this connection (or lack thereof) —which is typically not something one has any direct access to through reflection— provide (or fail to provide) one with justification for thinking one's belief is true? Put another way, if an agent possesses strong *internal* grounds for her belief, then the lack of an external (typically, causal) connection is *irrelevant* to the justificatory status of her belief, provided that such a connection is directly inaccessible to that agent. Internalists often try to motivate this objection through reference to the victims of demon worlds. I will defend reliabilism against the demon world victim objection. In doing so, I will offer a host of varied responses open to the reliabilist, clarifying the appropriately salient features present in the demon world scenario. Some of these responses simply deny the validity of the alleged intuitions about the thought experiment, while others accept these intuitions, but argue that they can be ultimately resolved within the reliabilist framework.

¹ My focus will be on the simple, ordinary beliefs of the demon world victims.

² I will be primarily focused on how internalist *justification* differs from externalist *justification*, but what I say regarding justification will apply *mutatis mutandis* to other epistemic concepts. Moreover, when talking about *knowledge*, there are, of course, Gettier considerations one needs to take into account.

As a brief general characterization, accessibilist versions of internalism hold that all of the factors constitutive of justification need to be “cognitively accessible” to the subject, where “cognitively accessible” refers to access of a certain kind.³ Acceptable kinds of access for non-inferential justification will include things such as access through reflection or introspection. Internalism requires that all of the features necessary for a belief to be epistemically justified for a given person be available to that person through one of these forms of access.⁴ Often, the

³ Internal state versions of internalism, such as Feldman and Conee’s *mentalism*, hold that the essential factors constitutive of justification will be mental states of the subject. However, there may be a worry on these views over whether they are truly *internalist* views, since reliabilism and other versions of externalism standardly employ mental states in their accounts of justification. Moreover, it is neither necessary nor sufficient that the features literally be internal states *qua* non-relational properties of the mind. Some forms of internalism allow that things other than mental states can be cognitively accessible, such as universals and sense data; some versions also allow the relevant states in question to be only potentially (i.e. either logically or lawfully possible) accessible (versus actual access). Fumerton, however, thinks that there are problems with construing internalism as either involving internal states or access conditions. Essentially, the claim is that it is not clear how to differentiate internal states from external states; moreover, access requirements are too weak or too strong, either allowing an externalist to hold them or generating regresses, respectively. His own view is that what really differentiates internalists from externalists is the naturalization of concepts and a commitment to either inferential internalism or inferential externalism. See Fumerton [1995], especially Chapter 3, for a more detailed discussion of this issue.

⁴ Moreover, a hybrid view will also be an externalist view, such as Alston’s [1988] ‘internalist externalism’. This sort of view claims that some of the factors required for justification must be cognitively accessible while others need not be. To be sure, there are alternative ways to distinguish between internalism and externalism other than those mentioned above, but many of these attempts are problematic. Briefly, for example, one might attempt to differentiate between internalism and externalism along the issue of ‘intellectual integrity’ or ‘responsibility’. On this score, one might construe internalism as the position which maintains that the subject pursues epistemic goals such as intellectual integrity—so long as we are following the best evidence available to us, we are justified. But, this characterization of internalism is problematic—externalism is not simply the denial of *this* position. Surely, externalists are also

requirement of accessibility is construed in terms of internal states. Externalism, on the other hand, is often viewed as simply a rejection of the thesis of internalism. A preliminary broad characterization of externalism is the view that some of the factors constitutive of justification can be external to the subject. More specifically, externalism allows that at least some of the features necessary for a belief to be epistemically justified for a given person need not be directly accessible to that person. Thus, a person's beliefs might be epistemically justified in virtue of facts or relations that are external to her subjective conception. Of course, externalism *can* allow reflectively accessible factors, but it denies that what affects justificatory status is restricted to them. According to reliabilism, reliability — a factor external to the subject — is constitutive of justification. Reliabilism is a form of externalism because whether a particular belief-forming process is reliable may not be something a subject has direct access to through reflection. According to a reliabilist foundational account of justification, there are two basic types of justification: the base is *non-inferentially* justified while the rest of our justification is *inferential*. The following is a formulation of the base and recursive clauses of a process reliabilist analysis of justification:⁵

S's belief that p at t is justified if and only if:

- (1) *S's belief that p at t results from a process that is (or a process token whose relevant type is) unconditionally reliable and belief-independent (or unconditionally reliable but belief-dependent⁶) [*non-inferential justification*].*

interested in epistemic goals, including even, intellectual integrity. In fact, as we will see shortly, many externalists include the notions of 'intellectual integrity' or 'responsibility' in their own analyses.

⁵ One might wish to add additional conditions to this justification condition when speaking about whether *S* *knows* that *p*. Goldman [1979], p. 13.

⁶ I argue elsewhere (J.W. Mulnix [2012], p. 268) that a process does not need to be belief-*independent* in order to be unconditionally reliable and justified

OR

(2) *S's belief that p at t results from a process that is (or a process token whose relevant type is) conditionally reliable, **and** the input beliefs into the conditionally reliable process are themselves justified [inferential justification].*

Looking closer at clause (1) above, a *belief-independent process* does not take any beliefs as inputs, while a *belief-dependent process* takes (at least some) beliefs as inputs. Unconditionally reliable processes may or may not take (at least some) beliefs as inputs. A process is *unconditionally reliable* if a sufficient proportion of its output beliefs would be true (and even if any input beliefs were false). Examining our inferential justification clause (2), a process is *conditionally reliable* when a sufficient proportion of its output beliefs are true *given that its input beliefs are true*. Conditionally reliable processes always take (at least some) beliefs as inputs. Accordingly, if most of the input beliefs to a conditionally reliable process are false, then it is less likely that the output beliefs will be true. Goldman explains: "A reasoning procedure cannot be expected to produce true belief if it is applied to false premises. And memory cannot be expected to yield a true belief if the original belief it attempts to retain is false."⁷ It is still very well possible that a majority of the output beliefs will be true, but at least, this is less likely. Note also that clause (2) specifies that not only must the process at work be reliable (conditionally), but the beliefs inputted

non-inferentially. My formulation of non-inferential knowledge then differs in some significant respects from the standard reliabilist formulation as given by Goldman, but is one, I believe, that best captures the reliabilist's commitments, and is something which has been to my knowledge previously overlooked. The key difference between my analysis and Goldman's is the notion of unconditional reliability in the base clause. I think Goldman has missed an important way in which even belief-dependent processes can be *unconditionally* reliable. See also J.W. Mulnix [2008].

⁷ Goldman [1979], p. 13.

to the reliable process must be *justified*. This is so because our analysis of justification is *recursive* —inferential justification is analyzed in terms of non-inferential justification. With this in mind, we are now ready to examine the demon world thought experiment.

1. Demon World Victims

Let us begin by imagining a possible world that is governed by an evil demon.⁸ This demon deceives all of the world's inhabitants, seeing to it that their most basic, ordinary beliefs are generally false, even though it seems to the demon world inhabitants that their beliefs are most obviously true. Now imagine that I have a twin in this possible demon world. The presumption in this thought experiment is that my twin and I share the exact same internal phenomenological characteristics (i.e. how things appear to each of us is identical), and we are both thinking that we see a table and adopt the belief that there is a table in front of us. Thus, these demon world victims have the same subjective internal bases for their beliefs that we do in our own world. But, it just so happens we are related to the world differently. I am standing in front of a table but my twin is not. Furthermore, my twin is subject to the machinations of the demon who sees to it that she hallucinates a table when there is no table, so that my twin *never* gets her perceptual beliefs right. I, on the other hand, am not being manipulated by a demon, and so, usually get things right. Thus, in the demon world, because of the demon's machinations, all of the demon victims' belief-forming processes are unreliable. For example, their simple perceptual, memorial, inductive (and even, perhaps, mathematical) beliefs are massively false, while our beliefs are generally true, even though the internal bases for the beliefs in the two worlds are the same.⁹

⁸ This thought experiment was initially formulated by Lehrer and Cohen [1983], though the specifics of the thought experiment in this paper vary from their original formulation.

⁹ An alternative version of the demon world problem can be found in Cohen [1983], according to which there are only two inhabitants in the demon

How would internalism and externalism evaluate this scenario? Focusing only on our simple ordinary beliefs, on any version of internalism our beliefs must have the same justificatory status — they are either justified or unjustified to the same degree. For example, on internal state versions of internalism, both my twin and I are in the same internal states, and so we are both equally justified in our belief about the table. And on accessibilist versions of internalism, we could argue that both my twin and I have direct access to the same features which would justify our belief in the table, namely, our perceptions. Conversely, according to externalism, I am justified while my twin presumably is not. More specifically, on *reliabilism*, we can point to a difference in the reliability of our processes, i.e. the process my twin is using is unreliable since the demon sees to it she never gets things right.

The key question that the internalist asks the reliabilist is: why, if my twin shares the same internal grounds for her beliefs as I do mine, would she not possess the same justification as I do? Or, is my twin unjustified in her beliefs because her processes are unreliable? But, if she is unjustified in her beliefs (because her beliefs are unreliably produced), in spite of sharing the same internal grounds as I do, then how can my beliefs be *justified*? The conclusion drawn by internalists is that justification must not simply be a matter of the reliability of one's belief-forming processes. For, because we can acknowledge that the demon victims do possess strong internal grounds for their beliefs even though their beliefs are unreliable, and because we want to attribute justification to these beliefs, reliability cannot be a necessary condition for justification. This objection challenges

world, one who reasons according to standard logic and the other who engages in confused reasoning. In this version, the demon ensures that both of their processes are unreliable, and so reliabilism would maintain that both reasoners' beliefs are unjustified. Cohen, however, argues that the reasoner who uses standard logic has justified beliefs. For the purposes of this paper, I have focused on the case in which one of the reasoners has reliably produced beliefs.

the core of reliabilism by maintaining that only internal factors matter to justification.

As I see it, the internalist relies on two key alleged intuitions of the demon world thought experiment when motivating this objection to reliabilism:

(I1) Whatever we say about the justification of our beliefs in our world, we need to say the same things about the justification of corresponding beliefs formed by the demon victims in their world.

(I2) The victims of the demon world are epistemically “responsible”; as such, their beliefs should be justified, even if it turns out that their faculties are not reliable.

The first alleged intuition is essentially the idea that the justificatory status of my beliefs and my twin’s beliefs must be the same, for the reason that the two worlds are relevantly similar. Fumerton [1995] has argued that the demon world thought experiment is actually *neutral* with respect to whether the demon world victims’ beliefs are justified. It only says that if ours are justified so must be theirs; if ours are unjustified, so must be theirs. For Fumerton, whatever we say about the two cases must be the same.

Now, since most believe that our world is *not* a demon world, they would regard our simple basic ordinary beliefs as justified.¹⁰ Thus, assuming for the purposes of this paper that we have ruled out the possibility that our simple, ordinary beliefs are unjustified, I1 ends up being equivalent to the claim that the beliefs are *justified* in both worlds. To deny the validity of I1 is to argue that the justificatory status diverges due to some salient dissimilarity among the two worlds, with the end result being that in one world the beliefs are justified, while in the other world

¹⁰ Though there are those, such as Fumerton, who deny that most of our beliefs about the external world are justified, even in a non-demon world.

they are not. Again, this is generally comparable to the claim that our beliefs are justified, while the demon world victims' beliefs are not justified.

The second principal alleged intuition is connected with the notion that 'justification' is fundamentally a *normative* concept associated with intellectual integrity, subjective rationality, and responsibility. For instance, some internalists maintain that a person whose overall set of beliefs is consistent and coherent is being epistemically "responsible" — the demon world victims are perfectly subjectively rational in their belief systems qua internal coherence in spite of a lack of connection to truth. Others might argue that "epistemic responsibility" is a function of whether a person can explicitly identify what she takes to be a good reason for holding a given belief. While the specifics may vary among different views, the general idea is that the demon world victims are forming their beliefs in an epistemically "responsible" way. This second intuition is also related, in part, to the further idea that 'justification' involves in some way or other "fulfilling one's duties" or being "epistemically praiseworthy" or "epistemically blameless." Of course, many standard internalists (and externalists) reject the idea that epistemic concepts are deontic concepts.¹¹ Even so, there is a sense in which one can understand justification as a normative concept without committing oneself to the claim that justification involves duties.¹² What is more, that the demon victims' are "doing the best they can" is a common judgment made about the thought experiment, and so it is

¹¹ See, for example, Alston [1985], Fumerton [1995], and Conee and Feldman [2001].

¹² It is important to note here that many contemporary externalists, such as Goldman for example, claim that justification is a *normative* concept. Technically speaking, on Goldman's reliabilism, justification is conceptually analyzed into non-evaluative or non-normative constituents. Goldman [1979] is very clear that the base clause of his account of justification must be non-circular and contain no epistemic terms. In fact, his base clause will only make reference to 'natural' properties. But, his account is still normative, he argues, in that justification, a normative property, supervenes upon these natural properties.

worth the reliabilist's time to consider and defuse this second alleged intuition as well. To deny the validity of I2 is to argue that 'epistemic justification' does not consist in, or at least is not exhausted by, *purely* internal grounds for believing something to be true, where these internal grounds imply that the agent is either being epistemically responsible or fulfilling her duties.

To be sure, these two intuitions are related. But, the acceptance of I1 still leaves it an open question whether one also accepts I2, since there are many views of epistemic justification (including internalist analyses) that do not link justification with normativity or "responsibility." For example, Richard Fumerton, an internalist, argues that there is no connection between praise and blame on the one hand and justification and rationality on the other. To make a judgment about justification is not to praise or blame *either* the agent or the belief:

To describe someone as being epistemically justified or rational, unjustified or irrational, in believing *P* is not in and of itself to make any moral or prudential claim about what he ought to believe. It is not to praise or blame the person for having the belief. It is not, I think, even to praise or criticize the belief. Of course, given certain values, it may be extremely important to us to have justified as opposed to unjustified beliefs.¹³

In evaluating these two intuitions, there are four clear routes one could take with respect to them:

- (1) Both I1 and I2 are true: We must say the same thing about the justificatory status of the beliefs in both worlds because the situations are similar; and

¹³ Fumerton [1995], pp. 19-20. Additionally, Plantinga [1993a; 1993b] thinks that deontic concepts (e.g. obligation, duty, what one ought to believe or refrain from believing) are not going to be the kinds of concepts one can plausibly use in the analysis of the third condition for knowledge. Instead, Plantinga uses "epistemic warrant" as whatever it is that must be added to true belief.

we ought to attribute justification in cases when agents are epistemically “responsible.”

(2) I1 is true, but I2 is false: We must say the same thing about the justificatory status of the beliefs in both worlds because the situations are similar; but justification is *not* connected to being epistemically “responsible.”

(3) I1 is false, but I2 is true: We need not say the same thing about the justificatory status of the beliefs in both worlds because the situations are dissimilar; but we ought to attribute justification in cases when agents are epistemically “responsible.”

(4) Both I1 and I2 are false: We need not say the same thing about the justificatory status of the beliefs in both worlds because the situations are dissimilar; and justification is *not* connected to being epistemically “responsible.”

We can glean a further insight by looking at these alternatives. Namely, *even if* one is able to deny the first intuition that the justificatory status of the two worlds must be the same, one would still need to say something about whether the demon world victims’ beliefs are justified or unjustified. And, if one maintains that the demon world victims’ beliefs are indeed *unjustified*, then one must explain why this is so despite the victims having formed their beliefs in an allegedly “responsible” way.

Actually, I think there is likely a fifth option as well, one that is closely aligned with option (4). Namely: I1 is false, and I2 is irrelevant.¹⁴ This fifth option does not specifically outright *deny*

¹⁴ Actually, I think there is likely also a sixth option, one connected with option (2). Namely, I2 is not necessarily false, but irrelevant. Option (2) endorses I1, which essentially amounts to the claim that the beliefs in both worlds are justified. In this case, one *need not* take a further stand about epistemic

that justification involves being epistemically “responsible” or subjectively rational (perhaps, these are included among the many conditions for justification, for instance), but argues that, in this case, it is irrelevant. It is irrelevant because once we determine that the situations are dissimilar, there is no further compulsion to either accept or reject I2. In a way, this option essentially argues that, once one denies the first alleged intuition, there remains nothing compelling about the second alleged intuition—it is beside the point. In that case, we would actually have (4a) and (4b):

(4a) I1 is false, and I2 is irrelevant: We need not say the same thing about the justificatory status of the beliefs in both worlds because the situations are dissimilar; and the question of whether the agents are epistemically “responsible” is beside the point.

(4b) Both I1 and I2 are false: We need not say the same thing about the justificatory status of the beliefs in both worlds because the situations are dissimilar; and justification is *not* connected to being epistemically “responsible.”

As further explanation, according to a basic straightforward reliabilism, reliability is necessary for justification. Thus, once we

responsibility, since the question of whether the demon world victims’ beliefs are justified has already been answered. An externalist analysis of this claim could argue that the demon world victims’ beliefs are, in fact, *justified* because their processes are *reliable*, and this is why we must say the same thing about the beliefs in both worlds. This option could be attractive for those reliabilists who want to base their notion of “epistemic responsibility” in some way on reliability. That is to say, a reliabilist need not outright reject the idea of “epistemic responsibility” when this notion is properly framed within externalist parameters. For our purposes, it is enough to keep in mind that one need not *outright* reject or endorse the idea of epistemic responsibility when holding the position that the demon world victims do have justified beliefs, for the reason, again, that the justification of their beliefs may or may not be a function of the demon victims’ acting “responsibly.”

determine that the demon world victims' processes are unreliable, the discussion is over — their beliefs cannot possibly be justified because reliability is necessary for justification. That is, at the moment that one of the necessary conditions for justification is not met (i.e. reliability), there is no further need to consider any other possible constituents of justification, and thus, nothing more needs to be said by way of I2.¹⁵ Option (4b), on the other hand, is a stronger “bite the bullet” version of (4a) for the reason that (4b) outright *denies* that ‘epistemic justification’ is *in any way* a function of being epistemically “responsible”.

Note also that this paper is focused on available *reliabilist* replies to the demon world thought experiment, and so, I will not focus on possible internalist responses. Instead, I would like to develop the various ways in which a reliabilist can either deny or accommodate these two intuitions *within reliabilism*. Again, to be sure, internalists strongly disagree with each other over the question of whether justification is inextricably linked with the notion of duty or praise and blame; and, even if justification is so connected, it is unclear whether justification is exhausted by these other notions. Nonetheless, once again, the claim that the demon victims' are acting “responsibly” is a common judgment made about the thought experiment, and to the extent that some internalists downplay this second intuition, all the better for reliabilists, I think. But, as we will see shortly, some reliabilists try to bring in these notions as indispensable constituents of *justification*, due in some measure, no doubt, to thought experiments such as this one. My own opinion is that all reliabilists should deny I2 as a feature of the concept of ‘epistemic justification’, for the only way to really accept I2 is to adopt some form of a virtue externalism or to accept a bifurcation of ‘justification’; however, in my view, these responses either appear to be *ad hoc*, or seem to

¹⁵ Perhaps another way of thinking about this could be that I2 supervenes on I1 — epistemic normativity supervenes on the natural property of reliability — and when we consider the thought experiment more carefully, we realize that I1 fails to obtain.

give too much ground to the internalist. More will be said about this later.¹⁶ In any case, my primary aim in this paper is to show that there are a host of varied responses open to the reliabilist that provide her with a way to defend reliabilism against the demon world objection.

2. Reliabilist Responses

Now, filling in these general options with the specifics of reliabilism gives us at least the following nine reliabilist responses, as we can separate out at least two different approaches within response (2), and at least four strategies contained by response (4a).¹⁷ After briefly characterizing each of these options below, I will follow with a more extended discussion:

(1) *Both I1 and I2 are true:*

[**Response 1**] Move from a Process Reliabilism to a version of a *Virtue Reliabilism*: the demon world victims' beliefs are justified because justification consists, in part, in exercising "intellectual virtues." And the two situations are similar because "intellectual virtues" are applied rigidly to all possible worlds (e.g. Goldman (1993) and (1999)).

(2) *I1 is true, but I2 is false (or irrelevant):*

¹⁶ Though, for an expanded defense of these claims, see my J.W. Mulnix, "Reliabilism, Demon Worlds, and Two Senses of Epistemic Evaluation" [in progress].

¹⁷ Though note that this list of nine responses is not meant to be exhaustive of every possible reliabilist response. In fact, one could add several more responses to this list, including the claim that while the demon world victims' processes are reliable, reliability is only necessary and not sufficient for justification; and, the demon world victims do not meet the other reliabilist conditions for justification, such as the lack of available relevant defeaters, etc.

[**Response 2**] The demon world victims' beliefs are justified because their beliefs are actually *reliably* produced (and not because they are epistemically "responsible"). The process-types are reliable in both worlds because perceptual beliefs carry with them *no ontological commitment*— or, we have no *de re* perceptual beliefs (e.g. phenomenalism).

[**Response 3**] Clarify the conception of *reliability* such that the demon world victims do have reliably produced beliefs: The demon world victims' beliefs are justified because their beliefs are actually *reliably* produced (and not because they are epistemically "responsible"). The process-types are reliable in both worlds because reliability is tethered to the actual world or a 'normal world' (e.g. Goldman (1979) or (1986)).

(3) *I1 is false, but I2 is true:*

[**Response 4**] *Bifurcate* the concept of *justification* into "strong" and "weak" components, according to which the demon world victims' beliefs are weakly justified: The two situations are dissimilar in that the demon world victims' processes are unreliable while our processes are reliable. Nonetheless, while the demon victims do not have reliably produced beliefs (and so, their beliefs are not "strongly" justified), the demon world victims do share something in common with us by way of their internal grounds, and thus, their beliefs are "weakly" justified because the demon world victims are epistemically "responsible" (e.g. Goldman (1988)).

(4a) *I1 is false, and I2 is irrelevant*: the demon world victims' beliefs are not justified because the two situations are dissimilar. The worlds are actually disanalogous. The demon world victims either have: (a) different phenomenal inputs; (b) unjustified input beliefs; (c) different process-types; or (d) divergent output beliefs. There is no need to take a stand on the question of whether the demon world victims are also epistemically "responsible":

[**Response 5**] The *inputs* to the demon world victims' processes are different from our inputs.

[**Response 6**] Perception is a belief-dependent process, and the demon world victims' *background beliefs* are unjustified while our background beliefs are justified.

[**Response 7**] The demon world victims' *processes* are different from our processes, or their processes are defective, while ours are not.

[**Response 8**] The demon world victims' *output beliefs* are different from our beliefs.

(4b) *Both I1 and I2 are false*:

[**Response 9**] "Bite the bullet": justification concerns beliefs and *not agents*. The demon world victims' beliefs are not justified because the two situations are dissimilar: the demon world victims' processes are unreliable while our processes are reliable. What is more, 'justification' is *not* connected to being epistemically "responsible."

Some of these responses deny or sidestep one or more of the alleged intuitions of the demon world thought experiment, while others accept one or more of the intuitions, but show how they can be accommodated within reliabilism. More specifically, in some cases, *the reliabilist explains away the thought experiment* by showing that it is a mistake to think of the situations as similar in any way, particularly when we focus only on the salient features of the scenario. In other cases, the reliabilist demonstrates that it is a mistake to think that the demon world victims' processes are unreliable. And, in a couple of cases, the reliabilist makes certain concessions to the internalist by trying to accommodate I2 through the use of either "intellectual virtues" or "weak justification." Also note that some of these strategies are interrelated, and in certain cases, the reliabilist can use some of them in conjunction with each other (though note that some responses are mutually exclusive). Now, in what follows, I will offer a more or less extended discussion of each these alternatives.

Response 1: This first type of response actually involves making some measured changes to one's definition of reliabilism, in order to accommodate both intuitions behind the thought experiment. Rather than advocate for a basic version of standard process reliabilism, one might, instead, move to a version of virtue epistemology, such as the reliabilist-brand of virtue epistemology of Goldman [1993], which he calls "virtue reliabilism."¹⁸ This theory claims that a belief is justified if it results from an "intellectual virtue" and unjustified if it results from an "intellectual vice." We inherit

¹⁸ See also Greco and Sosa. Very briefly, Sosa [1985] argues that his view eliminates the tension between internalism and externalism by introducing the notion of "epistemic virtue" and attributing justification to the agent rather than the belief; it is a person-based rather than belief-based approach to justification, and is modeled in many ways on virtue ethics. Though, perhaps not all versions of virtue epistemology will endorse I2 with respect to the demon world victims, depending on the analysis of "epistemic responsibility" (such as whether acting responsibly requires conscious endorsement of a rule or principle).

our list of virtues and vices from our social background. Basically, the concept of 'justified belief' is identified with "belief obtained through the exercise of intellectual virtues (excellences)":

The epistemic evaluator has a mentally stored set, or list, of cognitive virtues and vices. When asked to evaluate an actual or hypothetical case of belief, the evaluator considers the processes by which the belief was produced, and matches these against his list of virtues and vices. If the processes match virtues only, the belief is classified as justified. If the processes are matched partly with vices, the belief is categorized as unjustified....¹⁹

This kind of view addresses the demon world victim cases because an epistemic evaluator will match the demon world victims' perceptual processes to one of the items on her list of intellectual virtues and conclude the victims' beliefs to be justified.

Goldman [1999] further refines his view, stating that there are two stages of judgment on virtue reliabilism: the first stage is actual reliability (not what is judged or believed), and it is the root criterion of justification. Goldman calls this stage the *standard-selection stage* because "it involves the selection of approved epistemic standards, viz., the approved belief-forming processes or methods that confer epistemic warrant."²⁰ In other words, reliability is the basis for classifying which processes are "good" and which are "bad." The second stage is the *standard-deployment stage*, wherein members of the community *judge* whether these chosen standards regarding "good" and "bad" processes apply to *particular* beliefs (either actual or hypothetical). Thus, what is distinctive about this particular version of reliabilism is that judgments of justification occur at the second-stage, and so, these judgments do not appeal directly to reliability, but instead, to

¹⁹ Goldman [1993], p. 274-275.

²⁰ Goldman [1999], p. 11.

“pattern matching” of particular cases to stored prototypes of “good” and “bad” belief-generating processes.

Moreover, on Goldman’s view, the community members apply these standards “rigidly,” according to which each of the approved processes are judged to be justification-conferring in any possible world in which they operate. We can see how this approach easily tackles the evil demon problem:

Recalling the demon-world example, it may be asked whether the basis for process evaluation is performance in the actual world or performance in other possible worlds as well. Presumably, evaluators appeal to real-world track records in making their evaluations, but they may also tend to assume that these observed track records can be extrapolated to non-observed and non-actual cases. In general, it is doubtful that ordinary evaluators, who lack ways of thinking systematically about ‘possible worlds’, use any sharply defined, world-relativized basis for reliability assessments. The crucial point, however, is that evaluators do not directly apply reliability considerations to novel cases. According to the present theory, they do not say (or think): ‘Since perception is unreliable in the demon world, therefore perception-based beliefs in that world are unjustified’. Instead, they use pattern-matching to previously entrenched prototypes to arrive at an epistemic assessment of the target belief. With this understood, we have a form of reliabilism that has notable success in explaining evaluators’ judgments in ‘hard’ cases, that is, cases that are hard for simple reliabilism.²¹

²¹ Goldman [1998], p. 208.

Summarizing, this new variant of reliabilism is able to handle the demon world victim scenario because perceptual processes are on the list of virtues —they are deemed reliable— and so, regardless of the actual number of true or false beliefs produced by such a process in the demon world, the demon world victims' beliefs will be justified. But why, in the standard-deployment stage, are the standards applied rigidly, such that the “good” and “bad” processes are justification-conferring in all the possible worlds in which they exist? It is not clear that our processes have anything to do with the demon world victim's belief-forming processes, and vice versa. Why is the way they form beliefs in any way connected to the reliability of our belief-forming processes?

Moreover, it seems that this theory is able resolve the demon world problem only by creating a bigger “demon”: namely, if justification is now dependent on *judgments of perceived reliability*, has our externalist concept of justification been “internalized”? And what about the possibility of mass confusion about which of our processes are reliable, so that we all determine, for instance, that wishful thinking is reliable while inductive inference is unreliable? Of course, a reliabilist might respond that, despite the fundamental appeal to the subjective perspectives of evaluators, justification is still importantly dependent on *external* factors, and more importantly, is not something which need be directly cognitively accessible upon reflection —in that sense, it is not “internal.” On the strength of that reply, this response could remain a rejoinder at hand to the reliabilist.

Response 2: A second line of response, on the other hand, denies the second intuition, but agrees to the first intuition that the worlds are relevantly similar. More specifically, the response claims that we ought to attribute justification to the beliefs of the demon world victims because the demon world victims actually have *reliably* produced beliefs. And the second response explains that this is the case because perceptual beliefs carry with them no ontological commitment —we have no *de re* perceptual beliefs. There are many ways one could go about arguing for this claim,

though one particular way would be to adopt a phenomenalist account of sensations. For instance, on phenomenism, the content of our perceptual beliefs about physical objects are not the objects themselves, but the properties and relations of our sensory experience (e.g. sense-data). Thus, when a demon world victim believes that she sees a table, the content of her belief is simply that she has experienced (or is experiencing or will or would experience) certain sense-data. If this is true, then the demon's machinations will not matter, for all of the demon world victims' perceptual beliefs will be true.²² Once again, this response challenges one of the alleged intuitions of the demon world scenario by disputing the stipulation that the demon world victims would end up with massively false beliefs, while also preserving the intuition that the demon world victims' beliefs would be justified. The demon world victims' beliefs would be justified because their perceptual belief-forming process is reliable, yielding mostly true beliefs (likely taking the form of subjunctive conditionals that involve the sorts of sensations one would experience given the experiencing of other sensations).

Response 3: The third reliabilist strategy also argues that it is a mistake to think that the demon world victims' processes are unreliable in the first place. This third response is effective because it challenges one of the fundamental presuppositions of the demon world scenario: the reliability of a process is assessed

²² However, this response would not blunt the full force of the thought experiment, were it to include the stipulation that the demon world victims' beliefs about things other than material objects are false as well. Moreover, phenomenism faces a number of serious questions, such as why it is that our sense-data are obtainable in certain conditions, or how our sense-data are explainable only in terms of other sense-data rather than through presupposing the actual existence of the material objects, etc. There is not room here to discuss the particularities of phenomenism, though see especially Chisholm [1948] and Fumerton [1985] for more discussion and criticism. Obviously, one could also motivate this particular response without presupposing phenomenism, such as by adopting a version of a causal theory of perception that contains no ontological commitments.

by its performance in the world of the example. Without this assumption, it is not clear that the thought experiment implies the unjustifiedness of the demon world victims' beliefs. Instead, my twin's beliefs are justified because her belief-forming processes are, in fact, reliable after all. Essentially, this response claims that a reliabilist has at her disposal some alternate interpretations of reliability which render a different epistemic status on the demon world victims' beliefs.

Making the point transparent, *within reliabilism*, a reliabilist can counter the scenario without modifying anything in the analysis of 'justification', but instead, through re-clarifying the conception of 'reliability' that is the hallmark of reliabilism. For example, if what matters is *only* the reliability in the *actual* world, then use of the same (reliable) process-type as in the actual world (e.g. perception of close objects in good lighting) in any other world would also yield justified beliefs, *even in* a demon world. But, before we can assess whether this "actual world" construal of reliability (or one involving a "normal world") succeeds in resolving the demon world problem, we first need to take an in-depth look at the concept of 'reliability' in its various formulations within reliabilism.

Recall that reliabilism defines justified beliefs as those beliefs that are reliably produced.²³ As noted earlier, reliability is a function of processes, more specifically, a function of the *relevant type* of processes. Problems with distinguishing the relevant type

²³ Of course, for a given belief to be justified, being reliably produced alone may not be sufficient: one may need to factor in other possible considerations, such as the availability of relevant defeaters, etc. Also, one needs to delimit the relevant threshold for reliability. This has been interpreted differently by reliabilists. Some reliabilists, such as Kitcher [1980], require that certain processes (i.e. '*a priori*' processes) yield all true beliefs in order to be reliable (they are 'ultra-reliable', to be exact). But for the most part, and even for a majority of Kitcher's processes, yielding more true than false beliefs (i.e. over fifty percent), is sufficient for a process to be deemed 'reliable'.

of a belief-producing process are beyond the scope of this paper.²⁴ But, even if the relevant type of process is identified, there still remains the question as to how we are to define its *reliability*. Moreover, reliabilism is concerned with a process's ability to yield true beliefs in the actual world. That is to say, a reliabilist is interested in determining whether *our* beliefs are true or false, and so, she will be concerned with figuring out whether the belief-producing processes we actually have are reliable. But, it remains an open question how one understands the *reliability* of our processes, and whether other possible worlds will be appealed to in such considerations. Is reliability a function of actual track record or a function of lawful statistical probabilities? And, are we concerned with the process's reliability in this world or in other possible worlds? As will be clear shortly, not all reliabilists restrict their relevant definition of reliability to the operation of our belief-producing processes in the actual world.

It is also important to note that even though the reliability we are concerned with is the reliability of our *actual* belief-producing processes — those belief-producing processes that humans do, in fact, have in this world — we cannot simply use the *actual track record* of our belief-producing processes to determine if they are reliable. That is, we cannot just record how many output beliefs are true and how many are false, *because* some of our belief-producing processes may have *one* or *no* output beliefs. Now, of course, we have not settled the issue of how processes are to be individuated, but it is certainly *possible* that we have some belief-producing processes with only one output belief. In these cases, the process would be either 100% reliable or 100% unreliable.

Take, for example, the following scenario: Suppose that I have a belief-producing process that takes as its input my mental state

²⁴ This issue is a particularly troublesome worry, known as the Generality Problem. See especially Feldman [1985]; Conee and Feldman [1998]; and Feldman and Conee [2002] for a detailed formulation of the Generality Problem, and see Alston [1995]; Beebe [2004]; and Alston [2005] for some promising solutions to this problem.

of contemplating who will win the baseball game I am watching, and it outputs the belief that the Padres will win the game. Suppose also that this is the only time I, or anyone, ever utilize this process — it is the only time I or anyone else actually forms a belief in this manner. Now, in our scenario, it turns out that the Padres *do* win the game. Do we want to say, then, that because the reliability — understood as *actual track record* — for this belief-forming process is 100%, that this process is reliable? Certainly, this result should strike us as odd. Why do find we this result unintuitive? Basically, we imagine that were I to continue to form beliefs in this manner, I would start accumulating some false beliefs, and so, the opposite conclusion would eventually prove to be the case. It is easy for us to realize that defining reliability in terms of actual track record fails to consider what *would* be the case *were* the process to be repeated; in other words, the definition of reliability made no use of *counterfactuals*. Even if the reliability we are concerned with is the reliability in this world, we must make appeal to counterfactual situations for this world, and this most likely will appeal to possible worlds for the truth conditions for the counterfactuals that are true for this world.

Moreover, we omit a consideration of those possible worlds where the counterfactuals are not true (because there are different laws of nature, and so, the appropriate antecedent conditions are not satisfied), and instead, focus on the closest possible worlds (where the counterfactuals will be true). Thus, reliability is not defined according to what happens in *all* logically possible worlds. Accordingly, whether or not there may be possible worlds in which its inhabitants are subject to the machinations of a demon will not factor into the determination of the reliability for my belief-producing process, because, presumably, these possible worlds are far enough away from my world (and subject

to different laws which render the counterfactuals applying to this world false).²⁵

Note that on this aforementioned account of ‘reliability’, reliability is *relativized* to each possible world —reliability is understood in terms of what goes on in *that particular world*. More specifically, “for any cognitive process *C*, if *C* is reliable in possible world *W*, then any belief in *W* that results from *C* is justified.”²⁶ Using the counterfactual account given above, the reliability of the belief-producing processes for each world’s inhabitants would be determined by reference to the counterfactuals that are true of that world.²⁷ On this view, then, the determination of the reliability of *our* processes focuses on the counterfactuals which are true in *our* world. Even so, it is not altogether clear whether on a relativized conception of reliability we are to configure reliability in terms of either “statistical facts about the actual world, propensities that exist in the actual world, or counterfactuals about what would occur in the actual world were a belief-producing process used long enough.”²⁸

²⁵ Even if we could say that my process and the process of my demon world twin are of the same type, reliability is not calculated according to one general type: the unreliability of my twin’s process will not negatively affect the reliability of my process, in that the false beliefs produced by her perceptual process are not factored into calculating whether my perceptual process yields more true than false beliefs —I do not need to weigh in the false output beliefs of my twin’s perception process-type when figuring out whether my perception process-type is reliable.

²⁶ Goldman [1979], p. 16.

²⁷ This does not mean, however, that a given process must enjoy the same reliability over all possible worlds, unless one construes reliability by reference to a “normal world”, to be discussed shortly.

²⁸ Fumerton [1995], p. 114. Fumerton also notes that talk about propensities might even essentially reduce to a type of counterfactual account. An alternative concept of relativized reliability is offered by Alston [1995]: “A process type is reliable if and only if it would yield a high proportion of truths over a wide range of situations of the sort we typically encounter where ‘we’ refers to human beings on Earth” (p. 27). The next step would then be to figure out what situations are those under which a process “normally” or “typically”

And, while this account seems to get things correct with regard to us, nevertheless we may conclude that this account is incorrect if it gets us the “wrong” results in demon worlds. That is, if our intuitions tell us that the demon world victims possess justified beliefs, but according to the relativized interpretation of reliability, the demon world victims’ beliefs are not justified (because their processes turn out not to be reliable when considering only the counterfactuals that are true in their world), then, perhaps, we ought to rethink our definition of reliability.²⁹

Indeed, there are other diverse ways to understand reliability, ways which *would* allow us to accept I1 and the claim that the two worlds are relevantly analogous. These alternative interpretations of reliability (i.e. “actual world” reliability and “normal world” reliability) allege that the same concept of reliability applies to both worlds. On these accounts, the reliability of the demon world victims’ process-type is judged either according to the reliability of the operation of that process-type in the actual world or in a normal world; accordingly, the demon world victims’ beliefs *are justified*. Once again, however, the second intuition that this justification involves some notion of “responsibility” is rejected, or at the very least, irrelevant.

The first option is to judge the reliability of belief-forming processes *in all worlds* according to the operation of belief-forming processes in our *actual* world. This sort of move is proposed by Goldman [1979], when he suggests that, instead of a belief in possible world *W* being justified if and only if it results from a cognitive process *C* that is reliable in *W*, a belief in possible world *W* will be justified if and only if it results from a cognitive process

functions. But this is some kind of indication of how reliability would be determined without reference to possible worlds, and yet, still be relativized to environment.

²⁹ Or perhaps not. Instead, the reliabilist could revert to one or more of the other proposed nine reliabilist responses. Alternatively, the internalist might argue that this fact shows that reliability is irrelevant to one’s justification for a given belief.

that is reliable in *our* world. That is, a belief in possible world *W* is justified if it results from a process that is reliable in our *actual* world. This is supposed to solve the problem because the only reliability we are concerned with is the reliability of beliefs formed in our world, and presumably, our world is not manipulated by a demon. Nonetheless, this sort of move should strike us as troubling. Once again, it is unclear that our processes have anything to do with the demon world victims' processes — why is the way they form beliefs in any way connected to the reliability of our belief-forming processes? Assuming we can even identify a similar type across the two worlds, why should we think that the reliabilities are thusly so connected? This method of resolution utilizing the actual world seems unintuitive. What is more, this response appears to deny the fundamental externalist insight that justification is strongly linked with the having of true beliefs (for the reason that the demon world victims' beliefs would be justified in spite of being false), which is also unacceptable.

Perhaps, a more initially plausible alternative is to modify the concept of reliability from a relativized-to-environment tendency or an actual world tendency, instead, to a "normal world" tendency. On this view, the reliability of all worlds is tethered to the reliability in a "normal world." The rough idea is to understand a "normal world" along the lines of a possible world ordering procedure, along the lines of work done by Nozick.³⁰ Of course, Goldman does not provide a precise formulation of a "normal world," admitting that such a notion is "quite vague," but what he does say is the following:

We have a large set of common beliefs about the actual world: general beliefs about the sorts of objects, events, and changes that occur in it. We have beliefs about the kinds of things that, realistically, do and can happen. Our beliefs on this score generate what I shall call the set of normal

³⁰ Such as in Nozick [1981].

worlds. These are worlds consistent with our *general* beliefs about the actual world. (I emphasize ‘general’, since I count worlds with different particular episodes and individuals as normal.) Our concept of justification is constructed against the backdrop of such a set of normal worlds. My proposal is that, according to our ordinary conception of justifiedness, a rule system is right in any world *W* just in case it has a sufficiently high truth ratio in *normal worlds*. Rightness is rigidified for all worlds; but it is rigidifies as a function of reliability in normal worlds, not reliability in the actual world.³¹

The upshot to this account of reliability should be clear: if reliability is determined according to a process’s function in a *normal* world, then the demon world victim’s beliefs are justified in spite of operating unreliably in *their* world. That is, while the operation of the demon world victims’ belief-producing processes are unreliable in their demon world, the process-types are (likely) *reliable* in a “normal world.”

Yet, this approach faces formidable obstacles, not the least of which is the problem it creates for the inhabitants in our world, for *us*. Namely, how does our actual world relate to a “normal world”? It seems possible that there could be process-types that would be unreliable in a normal world, but which operate reliably in the actual world, or that there are process-types that are reliable in our world, but which could be *non-existent* in a normal world. It seems, then, that the normal world approach of defining reliability only solves the demon world problem by, once again, creating a bigger “demon”: it severs the connection of reliability with the actual world. What is more, this solution also has the strikingly odd consequence of removing reliabilism from securing a connection between justification and truth, which,

³¹ Goldman [1986], p. 107.

again, is one of the driving motivations underlying reliabilist accounts of justification. Fumerton explains: "Unless one has some independent reason for believing that this *is* a normal world, why would the concept of justification even be important to a truth seeker when it is now so obvious that having justified beliefs need not even make *probable* having true beliefs?"³²

Even more troubling, how does one determine exactly which processes are reliable in a normal world? Fumerton explains the problem as follows:

...Goldman suggests that we understand the reliability that determines the justification of someone's belief always by reference to reliability in *normal worlds*, where normal worlds are defined by reference to certain sorts of fundamental beliefs people have about *this* world. It is never all that clear which beliefs are to be included among the fundamental beliefs. They presumably include such presuppositions of our ordinary thought as that there is a past and that our memory has something to do with it, that there is an external world and that perception is causally affected by what happens in that world, and so on. We are *not* supposed to build so much into the characterization of a normal world that it becomes analytic that any process we take to be reliable is reliable, but it seems to me, nevertheless, that it may end up being analytic that certain fundamental processes we take to be reliable are reliable.³³

Thus, such a decision appears either to outright beg the question — by presupposing beliefs that are true of the actual world — or it makes the judgment an analytic truth, which is highly dubious.

³² Fumerton [1995], p. 115.

³³ *Ibid.*, p. 114.

Even Goldman himself no longer adheres to this “normal world” account of reliability, opting instead for his *virtue reliabilism*, which would be a variation on the first response. Of course, one need not move to a virtue reliabilism if one is unsatisfied with Goldman’s account of a “normal world”; one could simply return back to the actual world account of reliability offered earlier in this response, as proposed by Goldman [1979].³⁴ And because an actual world account of reliability does confer justification on the beliefs of the demon worlds victims, this response does provide the reliabilist with a way to accept the first intuition of the thought experiment while remaining consistent with reliabilism.

Response 4: The fourth reliabilist response explicitly endorses the second alleged intuition that the victims of the demon world are epistemically “responsible” by way of doing everything they can from their subjective perspective to act in the pursuit of truth, and so, it holds that their beliefs should be justified, even if it turns out that their belief-forming processes are not reliable. Thus, this view accommodates the second intuition by arguing that in the demon world case, the demon world victims do have something going for them epistemically, but it is not the typical reliabilist analysis of justification. Essentially, this line of response modifies the reliabilist analysis of ‘justification’ by bifurcating the concept of justification into two kinds: “strong” and “weak.” “Strong

³⁴ The key difference between the actual world account of reliability involved in Response 1 and the actual world account of reliability in Response 3 is that Response 1 involves making changes to the analysis of justification—thus transitioning away from a process reliabilism to a virtue reliabilism—by way of introducing two stages of judgment, while Response 3 sticks with simple process reliabilism, but simply tinkers with the relevant understanding of reliability. It is true that Response 1 also utilizes an actual world account of reliability, but within a different framework of justification: the reliability that serves as the basis for the selection of standards in the *standard-selection stage* is of those belief-forming processes operating in the *actual* world; these standards are then applied rigidly to all possible worlds (including demon worlds) in the *standard-deployment stage*. Accordingly, the actual world account of reliability in Response 1 confers justification on the beliefs of the demon world victims.

justification" is, in essence, the standard reliabilist concept of 'justification', according to which the reliability that determines whether or not a belief is *strongly justified* is the reliability of the process that produces the belief *in the world in which the belief is produced*.³⁵ "Weak justification," on the other hand, derives not from *reliability*, but from "epistemic blamelessness," where a subject is acting as best she can epistemically —she does not believe the process to be unreliable, nor has she any reliable way of discovering that it is unreliable. Weak justification allows for a distinct mode of assessment of the justificatory status of a belief, where what matters is whether the belief was formed in accordance with the believer's standards:

S's belief is weakly justified at the primary level if (1) the cognitive process that produces the belief is unreliable, but (2) S does not believe that that producing process is unreliable, and (3) S neither possesses, nor has available to him/her, a reliable way of telling that the process is unreliable. Finally, a further condition may be appropriate: (4) there is no process or method S *believes* to be reliable which, if used, would lead S to believe that the process is unreliable.³⁶

Weak justification, then, attributes a kind of justification to an agent's belief if this belief is formed in a responsible manner. In this sense, weak justification seems intimately connected to the "epistemic responsibility" of the agent. On this proposed solution, we can say of my twin that her beliefs, though *strongly* unjustified, are *weakly* justified. Consequently, this response accepts the second intuition that the demon world victims are acting "responsibly"

³⁵ The precise formulation is as follows: "S's belief that *p* is strongly justified if and only if (1) it is produced (or sustained) by a sufficiently reliable cognitive process, and (2) that the producing process is reliable is not undermined by S's cognitive state" (Goldman [1988], p. 59).

³⁶ *Ibid.*

by noting that we *do* share something in common with the demon world victims, namely, “weak justification.”³⁷ However, note that this fourth response also *denies* the *first* intuition that we must say the same thing about the justificatory status of the beliefs in both worlds. It denies this first intuition by acknowledging that there is a type of justification — “strong justification” — that I possess, but that my twin does not.

Still, it is unclear that such a maneuver is not *ad hoc*, as some argument must be made for why we need *two concepts* of justification. We should be wary of positing additional concepts when one could suffice. Moreover, some argument must also be given for the claim that one of these forms of justification is “strong” while the other is “weak”, and which one is which — why is the justification tied to reliability strong rather than weak, for example? More importantly, these two concepts of ‘justification’ appear wholly unrelated. Weak justification seems concerned with the epistemic blamelessness of an *agent*, while strong justification is connected to the reliable formation of the *belief*; in that case, why not use two distinct concepts to reflect this difference, one concept that signals only the evaluation of an agent rather than the belief (we might call this concept ‘subjective rationality’) — while the other concept designates only the evaluation of a *belief*, which is already reflected in our standard concept of ‘justification’. This line of thinking might be used to motivate Response 9, which we will consider later. But, there are many reliabilist responses that we have yet to consider. As with this fourth response, the remaining five responses will also deny the first intuition; but, in contrast with this fourth response, they will either ignore or reject the second intuition.

³⁷ Of course, an internalist might simply reply to this response by arguing that there is nothing “weak” at all about the justification possessed by the demon world victims — the demon world victims’ beliefs are either justified or unjustified (for internalists, presumably, their beliefs are justified), and there are not two distinct concepts of justification in play.

Response 5: One way to deny similarity between the two worlds is to argue that the *inputs* to the demon world victims' processes are different than the inputs to the processes when used in the actual world. This sort of response essentially denies the first alleged intuition that whatever we say about our world we need to say about the demon world, because it denies a similarity between the two worlds. Indeed, as I noted previously, it seems that one fundamental presupposition of the demon world thought experiment is that the two types of people have *indistinguishable* phenomenological experiences. Fumerton presents it this way:

Let us suppose that the 'software' in the minds of demon-world dwellers is precisely the same as the 'software' in our minds, and that the input and the output beliefs in both worlds are exactly the same, the input being that something like the data received through the five senses, the output being commonsense beliefs about the physical surroundings.³⁸

Figuring out whether my twin and I do indeed share indistinguishable phenomenological experiences naturally raises the question as to how to characterize the relevant inputs to our belief-forming processes.³⁹ Views differ over how to understand the relevant input: one might argue that the input "begins at the surface of the skin, or farther in at some point where conscious experience begins, or farther out in an external cause of the experience."⁴⁰ In the case of visual belief-formation, Alston argues that the relevant input is the conscious perceptual experience.⁴¹

³⁸ Fumerton [1995], p. 113.

³⁹ And this, of course, begs the question as to how to characterize the belief-producing process as well. See n. 24 in this paper.

⁴⁰ Conee and Feldman [1998], p. 27, n. 21.

⁴¹ Alston explains the reasoning for his position: "If the epistemic status of a belief is a function of the reliability of the process that generates the belief, it is the reliability of the psychological process that is crucial. Looking at perceptual

Certainly, if a reliabilist understands a belief-forming process in terms of inputs where the causal chains terminate in the mental states of the subject, then it is plausible that the demon world victims might have the same inputs as people in the actual world, since their conscious perceptual experiences would likely be indistinguishable from ours.⁴² However, there *may* be an important sense in which one could consider the inputs significantly *different*: Their perceptual inputs don't come from the same *causal origin*.⁴³ A maneuver the likes of this might be able to provide the reliabilist with some basis for a relevant distinction between the actual world and the demon world. It would then make sense that the demon world victims' processes might not be reliable because processes are reliable relative to certain inputs and conditions. Correlatively, it would also make sense that the demon world victims would have unjustified beliefs while our beliefs are justified.

Response 6: The sixth, related response also denies a resemblance between the two worlds. It does so by pointing to the *belief-dependent* nature of the belief-forming process in question. For instance, one standard view in the philosophy of perception is that perception is a belief-dependent process, taking in all sorts of background beliefs in addition to sensory inputs.

belief formation, no matter how exemplary the path of the light rays from the surface of the perceived object to the retina, and no matter how finely tuned the neural transformations involved in the pathway from the eye to the brain, if the belief is not formed on the basis of the conscious presentations (and/or its neural correlate) in a truth-conducive way, the belief will lack the epistemic desideratum that is stressed by reliabilism" (Alston [1995], p. 12).

⁴² Tracing the causal chains back any further threatens to trivially include the truth-maker (the physical object which is the subject of the belief), and hence, would make the process 100% reliable; but, this would violate the intuition that one needs to allow the possibility for a false justified belief.

⁴³ For instance, Tye [1995] argues that *mental content externalism* applies to the phenomenal content of our perceptual experiences as well, though this is quite controversial. Nonetheless, such a view could support this response that the perceptual inputs to the demon world victims' belief-forming processes are different. Also, see *Response 8* and n. 45 in this paper for further discussion.

These background beliefs might include such beliefs as 'My experience conforms to normal conditions'; 'Objects that look like x are x '; or 'I am not currently being deceived by an evil demon', etc. Now, while these background beliefs are true in the actual world, they are not true in the demon world. Correspondingly, it is likely these background beliefs are not justified in the demon world either (of course, one could try to focus on challenging this claim instead). Essentially, this is a variation on the idea that the demon world victims are working with different kinds of inputs to their belief-forming processes, but this time, the dissimilarity is not in the conscious perceptual experiences, but rather in the justificatory status of the associated background beliefs. Since these background beliefs — which are also inputs to the belief-producing process — are not justified, the demon world victims' output perceptual beliefs are accordingly not justified. Thus, once again, this allows us to make sense of the fact that the demon world victims would have unjustified beliefs while our beliefs are justified.

Though, note that this sixth response would be limited in that it applies only to those processes that can reasonably be construed as belief-dependent. For instance, if the demon sees to it that the demon world inhabitants believe in necessarily false basic axioms of geometry on the basis of counterfeit contemplations, then it seems as if the belief-forming process would be the same in our world as in the demon world, and so, this type of reply would not successfully explain away similarity in these non-inferential cases. But, one could then appeal to a difference in inputs in those cases. Or, one could simply deny that any justification is non-inferential and argue that all of our belief-forming processes are in some way or another belief-dependent; of course, my own foundationalist position is incompatible with this latter route. In any case, most often the demon world victim thought experiment focuses on the victims' perceptual processes.

Response 7: A reliabilist could deny correspondence between the two worlds by arguing that the perceptual belief-forming

process-type used in the demon world is different from the process-type used in the actual world. Again, this kind of response would first require a specification of what one means by a ‘process’. The two most likely candidates seem to be that processes are either mechanisms or functional states, the difference being that mechanisms do not appear to outright preclude the same process from taking different inputs as *relata*, while functional states are defined by the very inputs themselves, which would then make it impossible for the very same process-type to take different *relata* as inputs. Thus, depending on one’s account of processes, one might be able to make sense of the fact that my twin and I literally have different process-types, because we have different inputs (along the lines of either response five or six above).

Yet, even if we assume in the demon world and in our world the same process-type is used, we might argue that one of these processes is “defective” because belief-forming processes are designed only to handle certain kinds of inputs, namely those that arise from a particular *causal origin*. That is to say, the belief-forming process in the demon world is not equipped to accommodate certain inputs, namely “demon” inputs.⁴⁴ Thus, because the demon victims’ inputs are fabricated, the process outputs false beliefs, and this is the explanation for why the demon world victims’ process-type is not reliable, and hence, why their beliefs are not justified. Perhaps, an analogy might help here. Say, Claudia and Steve both have cars with the same type of engine. Claudia puts gasoline in her car and her engine works just fine. Steve, on the other hand, fills his car with orange juice and

⁴⁴ This could be understood either in a Plantinga-style “proper function” sense (see, for instance, Plantinga [1993a] and [1993b]), or in an evolutionary sense. For instance, the reliability of a particular process might have an evolutionary explanation, where “the process exists *as a result* of being truth-conducive. ...The existence of an evolutionary explanation (involving the past truth-conducive character of certain ways of arriving at belief) would be sufficient for a non-reflective (instinctive) reliable belief-producing process to yield apt belief” (Fumerton [1993], pp. 6-7 —though the context in which this statement is made is in describing Sosa’s views).

his car won't work. There is initially nothing wrong with Steve's engine—it is just that he puts in the wrong kind of *input*: one the engine was not designed to take in. Thus, now Steve's engine is "operationally defective," and his car cannot run. Returning to the demon case, what I have in mind here is, perhaps, there is nothing wrong with the belief-producing processes of the demon world inhabitants, but they are rendered unreliably "defective" given inappropriate inputs. Once again, the attempt here is to deny the alleged intuition that whatever we say about our world we need to say about the victims of the demon world.

Response 8: The eighth line of response yet again denies similarity among the two worlds. It does so by arguing that, in fact, the demon victims have different output *beliefs* than we do. Accordingly, we do not need to say the same thing about these two worlds because there are different beliefs in question. This strategy would utilize the insights of an externalism about meaning or mental content. For instance, on Kripke's [1972] and Putnam's [1975] causal theory of reference, the meaning and reference of natural kinds are determined in part by the surrounding environment. Thus, my duplicate twin's belief about a table in the demon world refers to something different than my belief about a table in the actual world. On mental content externalism, because external factors—whether these involve social communities and institutions (Burge [1979]), or the chemical composition of natural kinds (Putnam [1975])—are literally constitutive of belief-states, the demon world victims' beliefs would be different from our beliefs by virtue of being in a different environment with a different causal relation. That is, the demon world victims would have beliefs with different content. So, even if my twin and I are intrinsically alike in terms of conscious perceptual experiences, we hold different beliefs because beliefs do not supervene on intrinsic facts alone.⁴⁵ It follows, then, that because my twin and I

⁴⁵ Obviously, much more would need to be said here. For instance, it is not clear whether Burge's [1979; 1988] social externalism would work here since, presumably, the massive deception is not limited to a single case, but to all the

literally hold different beliefs, these beliefs need not be justified or unjustified to the same degree.

Response 9: The final reliabilist response, and my own preferred response, is to “bite the bullet” by denying *both* of the alleged intuitions of the thought experiment.⁴⁶ First, the reliabilist can deny that we need to say the same thing about the two worlds because justification concerns reliability rather than purely internal factors, and in one world the beliefs are reliably produced, while in the other world the beliefs are not. Second, the reliabilist can deny that we ought to attribute any justification to the demon world victims’ beliefs simply on the basis that they were being epistemically “responsible” for the reason that justification concerns beliefs and *not agents*.⁴⁷ On reliabilism, a *belief* is justified if it results from a reliable belief-producing process. Accordingly, the beliefs of the demon world victims’ are unjustified, and as such, the reliabilist analysis of justification gets things right. It is *irrelevant* that their beliefs are false through no fault of their own.

This view explains that the temptation to attribute justification to the demon world victims’ beliefs results from confounding one sense of epistemic evaluation with another, namely confounding

residents of the demon world. Moreover, one would need to say more by way of defending the presumed distinction between ‘narrow’ and ‘wide’ content (and the relation of externalism to each type), as well as the connection between the psychological contents of one’s belief-states and the linguistic contents of that-clauses involving belief ascriptions. Finally, it is not clear whether externalism applies to all beliefs or to only those beliefs involving ‘deferential concepts’ that involve fundamentally social elements. For example, externalism appears unable to handle the cases in which the demon world victims’ hold false beliefs about basic logical axioms. The literature surrounding externalism is vast, though Brown [2004] is a useful resource for further discussion of this issue.

⁴⁶ See my J.W. Mulnix, “Reliabilism, Demon Worlds, and Two Senses of Epistemic Evaluation” (in progress) for a defense of this position. Versions of this position also seem to be defended in Bach [1983] and Engel [1992].

⁴⁷ The view that the concept of the justification of a belief should not be tied to the blamelessness of the agent has been advocated by others before. See Pryor [2001] for a nice overview of the discussion, including both internalists and externalists who deny any form of justification tied to deontic concepts.

the sense of epistemic evaluation concerned with agents (along the lines of 'subjective rationality' or some related concept) with the sense of epistemic evaluation concerned with beliefs ('justification').⁴⁸ Yet, these two distinct senses do not represent a bifurcation of justification in the way Goldman advocated for above in Response 4. Rather, I think while there are two senses of epistemic evaluation, there is only one straightforward notion of *justification*, that which concerns *beliefs*. Instead, the two senses of epistemic evaluation correlate to the distinction between the evaluation of *agents* –in terms of their subjective rationality and not in any moral sense– and the evaluation of *beliefs*. That is, there is a sense in which one might believe what one is not justified in believing, and hence, be unjustified in one's belief; but yet, because this belief makes sense within one's belief system, and so one is adhering to some sort of subjective rationality, this may render one either epistemically "responsible" or "blameless."⁴⁹ Nonetheless, justification *proper* concerns the evaluation of beliefs, and so we should not attribute justification to the beliefs of the demon world victims, even if there is *some* sense in which they are epistemically blameless.

It might be helpful to draw upon an analogy with a utilitarian ethics here.⁵⁰ According to certain forms of utilitarianism, it is

⁴⁸ Again, not all reliabilists will be comfortable separating out the evaluation of beliefs apart from the evaluation of agents, especially on many forms of virtue reliabilism. For a competing view, see Greco [2006]. Though again, it should be emphasized that there are many *internalists* who also advance the view that justification concerns beliefs and not agents, such as Fumerton [1995].

⁴⁹ And the reverse is also possible: Failing to believe what one is justified in believing.

⁵⁰ I think the analogy between reliabilism and utilitarianism is particularly apt here, for the reason that both theories share an affinity by way of their naturalism based in externalist causal considerations. To be sure, this analogy will not work with many other ethical views, such as Kantianism, where it is not possible to separate out an evaluation of a moral action from the agent. And of course this makes sense given that I am denying that justification is a deontic concept. To the extent that one is resistant to utilitarianism one will find my

simply a mistake to think that there is any necessary connection that holds between the moral evaluation of an agent and the moral evaluation of an action. Accordingly, it is quite common for us to say things like:

She did/would have done the right action, but she was not praiseworthy/would not be praiseworthy (or perhaps even, would be blameworthy) for doing so.

She did/would have done the wrong action, but she was blameless/would be blameless (or perhaps even, would be praiseworthy) for doing so.

In many of these cases, the subjects either have good reasons to believe the opposite of what turns out being the case, or the subjects lack good reasons to believe what turns out to be the case.

In response to the question — “When are moral agents deserving of praise or blame?” — it is natural to respond that they are deserving of praise insofar as they have done the right thing, and conversely, are blameworthy when they have behaved wrongly. Of course, this is not the only possible answer to the question. In fact, one of John Stuart Mill’s valuable insights is that a person’s character may not be tied to the performance of right acts. That is, whether a person does the right thing in a particular case may or may not be relevant to whether that person ought to be praised or blamed for their action. Given Mill’s views about knowing the future consequences of our present acts, it is quite possible for an agent to do the wrong thing, in the sense that the action she chooses results in a lower utility than alternatives, and yet, be such that she is entitled to praise (or at least, not blame). In other words, the moral agent is not judged according to her having acted rightly or wrongly, but instead, on other grounds.

analogy less than compelling. Nevertheless, for a nice thorough account for why a utilitarian might be inclined to draw out a separation between the evaluation of agents and actions, see M.J. Mulnix [unpublished manuscript].

Mill is not concerned to explicate how we ought to evaluate moral agents, but he does suggest that there may be several factors. It might be whether the person had rational beliefs concerning the likelihood of their action to promote general welfare, but it might *instead* be whether the person had good *intentions* or acted virtuously. Whatever the basis for the moral evaluation of agents, it is not part of the analysis of right action.

On the simplest form of act utilitarianism, crudely put, an action is right or wrong depending on whether it results in the best consequences relative to alternative possible actions. Evaluating the moral rightness or wrongness of an action involves no reference to the agent's intentions or motives or character; moreover, it is the *action* that is the subject of evaluation and *not the agent* doing the action. The scope of the Utility Principle is limited only to the evaluation of moral acts and does not include the moral character of agents. Indeed, note Mill's own discussion of this issue in Chapter Two of *Utilitarianism*:

He who saves a fellow creature from drowning does what is morally right, whether his motive be duty, or the hope of being paid for his trouble.... But the motive, that is, the feeling which makes him will so to do, when it makes no difference in that act, makes none in the morality: though it makes a great difference in our moral estimation of the agent, especially if it indicates a good or a bad habitual *disposition* — a bent of character from which useful or from which hurtful actions are likely to arise.⁵¹

Right action is right, according to Mill, regardless of the reasons for acting. In fact, Mill's own example of the man who saves someone from drowning in the hopes of a reward shows

⁵¹ Mill [1861], Chapter 2, paragraph 19, including footnote.

that he thought a person can do the right thing even if they do so for quite insidious reasons. Mill elaborates:

These considerations [about praiseworthiness and blameworthiness] are relevant, not to the estimation of actions, but of persons; and there is nothing in the utilitarian theory inconsistent with the fact that there are other things which interest us in persons besides the rightness and wrongness of their actions. ...[Utilitarians] are also aware that a right action does not necessarily indicate a virtuous character, and that actions which are blamable often proceed from qualities entitled to praise. When this is apparent in any particular case, it modifies their estimation, not certainly of the act, but of the agent.⁵²

It seems clear that the rightness of an action should not be identified, even partially, with the agent's having justified beliefs regarding the likely results of their behavior. Although the having of such beliefs as a motive for acting the way we do might factor into others' estimation of us as "good" or "bad" persons (among other considerations), it is not relevant to the moral status of the actions we perform. Mill was not concerned to explain how it is that we should evaluate character, though he does hint that motives, reasons, and sincerity of intentions may play some role.

Let us motivate these ideas with an example: For instance, on actual consequence versions of utilitarianism (where the consequences that matter are the actual outcomes as opposed to probable or possible outcomes), an agent who enters the mall and shoots the first person she sees—which person happens to be a terrorist about to detonate a bomb that would kill everyone—undertakes the *right* action, despite having no intention whatsoever to save any lives. The best consequences were produced through

⁵² *Ibid.*, paragraph 20.

her action. On the actual consequence version of utilitarianism, her action promoted the greatest utility; nonetheless if one were asked whether or not the agent was “praiseworthy,” the answer would be less clear, and likely, negative.

In this and other cases, the moral status of the action is conceptually distinct from any moral evaluation of the agent. Fumerton puts the point this way:

Those moral philosophers who support an actual consequence act conception of right action nearly always try to ‘soften’ the view and make it more palatable by distinguishing sharply the concept of a good person from the concept of a person who (usually) does the right thing. The good person, it will be argued, should be understood in terms of her *intentions*. ... Praise and criticism, reward and punishment, the argument continues, is much more appropriately tied to evaluation of a person’s motives in acting rather than the rightness or wrongness of her action.⁵³

Moreover, in many cases, one is not able to foresee the actual consequences of a situation, so that one might be perfectly justified in performing a particular action that nonetheless turns out to have consequences other than those intended. In other words, a moral agent has no access to future states of affairs, including those that follow from her presently behaving in certain ways. Thus, it seems obvious that either utilitarianism demands the impossible from moral agents, or else there needs to be some way of evaluating the choices a person makes other than the rightness or wrongness of her action.

Now, why might we think that this analogy with utilitarian ethics is useful? Well, it does seem as if we can draw a connection between epistemic rationality and normative rationality. Indeed,

⁵³ Fumerton [1990], p. 105.

both domains deal with “ought” claims, so perhaps, they share a similar sort of structure. That is, just as the questions of what one ought to do and what is right to do are separate questions, so too, are the questions of what one ought to believe and what one is justified in believing.⁵⁴

Using this analogy, it is unclear why my twin’s “blamelessness” should at all enter into our analysis of the justification of her beliefs. Like certain views in ethics which mistakenly conceptually link the moral status of an action with the moral character of an agent, epistemic internalists, too, mistakenly connect the justification of a belief to the epistemic “responsibility” of the subject. Reliability is a theory of justification, which concerns the epistemic status of *beliefs*, not of agents; as such, the demon world victims, try as they might, have massively unjustified beliefs. Simply put, there is nothing “good” about the demon world victims’ *beliefs*, and though it is through no fault of their own, they do not have justified *beliefs*. It follows from these considerations that it is possible for a subject to be “epistemically praiseworthy” with mostly unjustified beliefs and “epistemically blameworthy” with mostly justified beliefs.

3. Concluding Remarks

As a final point, perhaps the most natural response to offer the internalist is simply to say: if the result that the demon world victims’ beliefs are unjustified strikes you as odd, the oddity of the outcome is not a function of the reliabilist account, but of the possible worlds we are being asked to consider. In other words, when considering odd possible worlds in which there are typically normal, and yet unreliable, belief-producing processes, it should not alarm us that these typically justified beliefs get classified as unjustified. Again, the oddity or lack of “intuitiveness” present is due to the odd type of world we are being asked to imagine, and

⁵⁴ See Fumerton [1990] for a compelling argument that the epistemic and moral “ought” share a similar structure.

hence, does not necessarily represent a bona-fide counterexample to the reliabilist position. This raises an interesting question about the role of intuitions in counterexamples to proposed definitional analytic truths about justification, but there is not room here to discuss this particular issue. Nonetheless, we ought to be cautious about relying too heavily upon certain intuitions about other possible worlds that heavily rely on intuitions about our own world.

In conclusion, I have canvassed a number of readily available reliabilist responses to the demon world objection. Once the appropriately salient features of the demon world scenario were clarified, it became clear how certain intuitions could either be accommodated or discarded within the reliabilist framework. While some of these reliabilist responses simply denied the validity of the alleged intuitions about the thought experiment, others accepted these intuitions, but argued that they could be ultimately resolved within the reliabilist framework. Finally, while many of these options do require further defense and elaboration, I have at least shown that there is a multitude of various ways that a reliabilist could address the demon world objection and maintain that reliability continues to remain the *hallmark* of justification.⁵⁵

References

- Alston, W. (1985). "Concepts of Epistemic Justification", *Monist* 68: 57-89.
- . (1988). "An Internalist Externalism", *Synthese* 74: 265-283.
- . (1995). "How to Think About Reliability", *Philosophical Topics* 23: 1-29.
- . (2005). *Beyond 'Justification': Dimensions of Epistemic Evaluation*. Cornell UP, Ithaca.

⁵⁵ I would like to express thanks to my anonymous reviewers for their helpful feedback.

Bach, K. (1985). "A Rationale for Reliabilism", *Monist* 68: 246-263.

Beebe, J. (2004). "The Generality Problem, Statistical Relevance, and the Tri-Level Hypothesis", *Nous* 38: 177-195.

BonJour, L. (1980). "Externalists Theories of Empirical Knowledge", *Midwest Studies in Philosophy* 5: 53-74.

Brown, J. (2004). *Anti-Individualism and Knowledge*. MIT Press, Cambridge.

Burge, T. (1979). "Individualism and the Mental", *Midwest Studies in Philosophy* 94: 73-122.

—. (1988). "Individualism and Self-Knowledge", *Journal of Philosophy* 85: 649-63.

Chisholm, R. (1948). "The Problem of Empiricism", *Journal of Philosophy* 45: 512-517.

Cohen, S. (1984). "Justification and Truth", *Philosophical Studies* 46: 279-295.

Conee, E., and R. Feldman. (1998). "The Generality Problem for Reliabilism", *Philosophical Studies* 89: 1-29.

—. (2001). "Internalism defended". In H. Kornblith (ed.), *Epistemology: Internalism and Externalism*. Blackwell, Malden, 231-260.

Engel, M. (1992). "Personal and Doxastic Justification in Epistemology", *Philosophical Studies* 67: 133-150.

Feldman, R. (1985). "Reliability and Justification", *Monist* 68: 159-174.

Feldman, R., and E. Conee. (2002). "Typing Problems", *Philosophy and Phenomenological Research* 65: 98-105.

Fumerton, R. (1985). *Metaphysical and Epistemological Problems of Perception*. U. of Nebraska Press, Lincoln.

—. (1990). *Reason and Morality*. Cornell UP, Ithaca.

—. (1995). *Metaepistemology and Skepticism*. Rowman & Littlefield, Maryland.

Goldman, A. I. (1967). "A Causal Theory of Knowing", *Journal of Philosophy* 64: 355-372.

- . (1979). "Reliabilism: What is Justified Belief?" In G.S. Pappas (ed.), *Justification and Knowledge*. D. Reidel, Dordrecht, 1-23.
- . (1986). *Epistemology and Cognition*. Harvard UP, Cambridge.
- . (1988). "Strong and Weak Justification", *Philosophical Perspectives* 2: 51-69.
- . (1993). "Epistemic Folkways and Scientific Epistemology", *Philosophical Issues* 3:271-285.
- . (1998). "Reliabilism". In E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Routledge, London, 204-208.
- . (1999). "A Priori Warrant and Naturalistic Epistemology", *Philosophical Perspectives* 13: 1-28.
- Greco, J. (2006). "Virtue, Luck and the Pyrrhonian Problematic", *Philosophical Studies* 130: 9-34.
- Kitcher, P. (1980). "A Priori Knowledge", *Philosophical Review* 89: 3-23.
- Kripke, S. (1972). *Naming and Necessity*. Harvard UP, Cambridge.
- Lehrer, K., and S. Cohen. (1983). "Justification, Truth, and Coherence", *Synthese* 55: 191-207.
- Mill, J.S. (1861). *Utilitarianism*. In J.B. Robson (ed.), *The Collected Works of John Stuart Mill*, (1963-1991). Univ. of Toronto Press, Toronto; Routledge and Kegan Paul, London.
- Mulnix, J.W. (2008). "Reliabilism, Intuition, and Mathematical Knowledge", *Filozofia* 62: 715-723.
- . (2012). "Rethinking the A Priori/A Posteriori Distinction", *Logos & Episteme* 3: 261-277.
- . (in progress). "Reliabilism, Demon Worlds, and Two Senses of Epistemic Evaluation".
- Mulnix, M.J. (unpublished manuscript). "Act Utilitarianism and Moral Character".
- Nozick, R. (1981). *Philosophical Explanations*. Harvard UP, Cambridge.
- Plantinga, A. (1993a). *Warrant: The Current Debate*. Oxford UP, Oxford.

—. (1993b). *Warrant and Proper Function*. Oxford UP, Oxford.

Putnam, H. (1975). "The Meaning of 'Meaning'". In *Philosophical Papers, Vol. II: Mind, Language, and Reality*. Cambridge UP, Cambridge, 215-271.

Pryor, J. (2001). "Highlights of Recent Epistemology", *British Journal for the Philosophy of Science* 52: 95-124.

Sosa, E. (1985). "Knowledge and Intellectual Virtue", *Monist* 68: 226-245.

—. (1991). *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge UP, Cambridge.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press, Cambridge.