

“Damming the swamping problem, reliably”

Jared Bates, Hanover College<sup>1</sup>

*dialectica* forthcoming

Abstract: The swamping problem is the problem of explaining why reliabilist knowledge (reliable true belief) has greater value than mere true belief. Swamping problem advocates see the lack of a solution to the swamping problem (i.e., the lack of a value-difference between reliabilist knowledge and mere true belief) as grounds for rejecting reliabilism. My aims here are (i) to specify clear requirements for a solution to the swamping problem that are as congenial to reliabilism’s critics as possible, (ii) to clear away various existing reliabilist solutions on the basis of these requirements, and (iii) to present a reliabilist solution that succeeds in meeting all of them. To meet all the requirements, my solution develops a more nuanced understanding of the epistemic end than is currently discussed, and with it a novel way of individuating beliefs. I close with a brief discussion of the question whether reliabilism’s critics might impose further demands which reliabilism cannot possibly meet.

The value problem in epistemology is the problem of explaining the value of knowledge—in particular, why knowledge has greater value than mere true belief. When the value problem is turned into a refutation of simple process reliabilism (hereafter, “reliabilism”), it is called the swamping problem.<sup>2</sup> Reliabilism holds that knowledge *just is* reliably formed true belief. The swamping problem asserts that while reliability may have value, its value is derived from the value of truth, so a belief that is both reliably formed and true has no more value than a belief that is merely true. The value of truth is said to “swamp” the value of reliability, and reliabilism is rejected for its inability to solve the value problem. My aims here are (i) to specify clear requirements for a solution to the swamping problem that are as congenial to reliabilism’s critics as possible, (ii) to clear away various existing reliabilist solutions on the basis of these requirements, and (iii) to present a reliabilist solution that succeeds in meeting them. To meet all the requirements, my solution develops a more nuanced understanding of the

---

<sup>1</sup> Department of Philosophy, Hanover College, Hanover, Indiana 47243, USA; Email: bates@hanover.edu.

<sup>2</sup> The objection appears in its most complete form in Ward Jones (1997, 425), Richard Swinburne, (1998, 58); Linda Zagzebski (1996, 301-302; 2000, 113-114; 2003, 13-14), and Jonathan Kvanvig (2003, 45-50). The list is not only (roughly) chronological; it is also (roughly) in increasing order of detail given to formulating the swamping problem, culminating in Kvanvig’s presentation, where it gets its name (2003, 45).

epistemic end than is currently discussed, and with it a novel way of individuating beliefs. The solution I present raises the question whether reliabilism's critics might impose further demands which reliabilism cannot possibly meet. I will close with a brief discussion of this question.

1. The swamping problem and what it demands.

The swamping problem begins with the thesis that the value of knowledge is greater than the value of mere true belief. Call this thesis *the greater value of knowledge*:

$$(GVK) \quad V(K) > V(MTB)$$

For purposes of reductio, assume *reliabilism*, namely, the thesis that knowledge just is reliably formed true belief:

$$(REL) \quad K = RTB$$

By substitutivity, GVK and REL entail *the greater value of reliabilist knowledge*:

$$(GVRK) \quad V(RTB) > V(MTB)$$

However, advocates of the swamping problem argue that GVRK is false. They do so by noting that reliability itself has only *derived* value, specifically that its value is derived from the value independently possessed by true belief. To see this, suppose that our epistemic end is to believe truly and avoid error. A state or condition has epistemic value if it stands in an appropriate relation to that end. A state or condition could, for instance, *constitute* that end. This is the case with true belief. Alternatively, a state or condition could be *instrumental* toward that end (i.e., could *cause* or *increase the chances of* that end's satisfaction). This is the case with being reliably formed, since reliable processes produce a higher ratio of true belief than unreliable ones produce. But on this approach the value of reliability is entirely derived from the value of true belief. Both reliable true beliefs and mere true beliefs are true beliefs, and so they equally constitute the epistemic end. Linda Zagzebski's popular espresso example brings

this point into relief.<sup>3</sup> Suppose the sole good-making feature of an espresso is that it is delicious. Then, intuitively, two cups of espresso that are equally delicious are equally good, even if one was made by a reliable espresso maker and the other by an unreliable one. So a good espresso is not made better by being reliably produced. Analogously, reliability doesn't *enhance* the value of true belief. Call this *the swamping thesis*:

(ST)  $V(\text{RTB}) \not\approx V(\text{MTB})$

ST is just the denial of GVRK, so, by reductio, REL is false.

In laying out the requirements on a reliabilist solution to the swamping problem, I want to concede as much as possible to reliabilism's critics. The following requirements set the terms of the problem:

1. A reliabilist solution must acknowledge the logical conflict between GVK, REL and ST (and not merely the *appearance* of conflict). One might try to deny the conflict by arguing that the value context,  $V(x)$ , is referentially opaque, and in such contexts substitutivity fails. I know of no solution that explicitly takes this approach. Reliabilism's critics assume an objective conception of value that permits substitutivity, and this assumption will be granted.
2. A reliabilist solution must preserve both GVK and REL. A response that denies REL may be a solution, but it isn't a reliabilist one. And a response that denies GVK simply fails to address the problem even on its most basic terms.
3. From (1) and (2), it follows that a reliabilist solution must entail the denial of ST. But more needs to be said, for the denial of ST is deducible from GVK and REL. Taking the swamping problem seriously requires reliabilists to *explain* GVRK, not simply by deducing it from GVK and REL, but by accounting for how reliability enhances the value of a true belief.
4. Moreover, the value added by reliability to a true belief must be *epistemic value*, and not some other sort of value. If we earned a nickel for every reliable belief we formed, and nothing for unreliable beliefs, then reliable true beliefs would have greater value than mere true beliefs, but not greater epistemic value. A reliabilist solution must explain how reliability enhances the *epistemic* value of a true belief.
5. The (epistemic) value added by reliability must accrue to the belief itself, and not to some more inclusive state or condition of the epistemic agent who holds the belief.

---

<sup>3</sup> See Zagzebski (2000, 113; 2003, 13). Richard Swinburne (1998, 58) makes essentially the same point about the good-making features of a desk and reliable furniture factories.

Reliabilism's critics have saddled reliabilism with a "machine-product model of belief" (Zagzebski 2003, 14), on which beliefs are conceived as outputs of cognitive processes. A reliabilist solution must explain how reliability imparts epistemic value to true beliefs so conceived.

6. And finally, in explaining *GVRK*, it will not do to point out that people tend to *attribute* greater value to reliabilist knowledge than to mere true beliefs (Kvanvig 2010, 103). What is required, once again, is to explain why instances of reliabilist knowledge have greater epistemic value than mere true beliefs. If I want to know why Toyotas are (alleged to be) better than other cars, it will not do to point out to me that people tend to pay more for used Toyotas – a kind of value-attributational behavior – than they will pay for other used cars. Even granting that such a tendency exists, the original question remains unanswered: What makes Toyotas better?

Not all reliabilist responses satisfy all these conditions. Not that violations are always unmotivated. But it is worth investigating how far reliabilism can go toward solving the swamping problem on its critics' terms. As I will show in section 3, reliabilism can adequately solve the swamping problem on the standards assumed by its critics. As will become clear in that section, explaining how reliability enhances the epistemic value of a belief requires revisions to both how we think about our epistemic ends *and* how we individuate beliefs. The section presents a richer, more nuanced understanding of these issues than the current debate presupposes.

## 2. Reliabilist solutions that fall short of the requirements.

Just how austere the above standards are can be seen by surveying the many existing reliabilist solutions that fall short of them. To begin, many assume that reliabilists have no choice but to escape the swamping problem by evacuating to the higher ground of virtue. Virtue reliabilists like Ernest Sosa (2007, 87) explicitly think that reliabilism succumbs to the swamping problem. Sosa abandons reliabilism by identifying knowledge with a certain sort of epistemic "performance" (2007, 23) which must be evaluated in virtue-theoretic terms and not merely on reliabilist conditions.<sup>4</sup> Since they reject REL, virtue reliabilist accounts violate condition (2) of the swamping problem and can be set aside.

---

<sup>4</sup> John Greco (2002; 2003) also gives a virtue reliabilist account of GVK. For an argument that not even virtue reliabilism can explain GVK, see Berit Brogaard (2006).

Wayne Riggs (2002) offers a solution to the swamping problem that cleaves nearer to the simple version of reliabilism we are considering here. However, like virtue reliabilists, his account of GVRK focuses on the epistemic credit agents earn by forming reliable true beliefs. In his words:

The value added by the non-accidentality of a true belief [i.e., by a true belief's being reliably formed] is value that *is added to the achievement of the person*. When a true belief is achieved non-accidentally, the person deserves epistemic credit for this that she would not be due had she only accidentally happened upon a true belief. This extra value is neither instrumentally derived from the value of having true beliefs, *nor is it a value that accrues to the belief itself*. (2002, 93, my emphasis)

Riggs later adds that there is no value-difference between reliable true belief and mere true belief, when the evaluation is at the level of the beliefs themselves (2002, 94). This account of GVRK, while it leaves REL intact, violates condition (5) by locating the value-difference between reliabilist knowledge and mere true belief at the level of the agent and not the beliefs themselves.

There are other reliabilist responses to the swamping problem guilty of violating condition (5). Erik Olsson (2007; 2009; Goldman and Olsson 2009) has developed an approach to the swamping problem he calls the conditional probability solution. He gives two versions, and I'll consider both presently. According to the first version, the probability of forming more true beliefs in the future is greater conditional on possessing a reliable true belief than on possessing a mere true belief (2007, 345; Goldman and Olsson 2009, 28). This is because possessing a reliable true belief entails possessing reliable processes, a state of affairs not required of possessing a mere true belief, and reliable processes are more likely to produce future true beliefs. A similar strategy is adopted by Justin McBrayer (2007), which he terms "the entailment strategy." However, this approach violates condition (5) of the swamping problem, which requires locating the value-difference at the level of beliefs themselves. The states of affairs that fall under epistemic appraisal for Olsson and McBrayer are broader than the beliefs themselves; in the case of reliable true beliefs, they include having and using reliable processes. Olsson explicitly states that the value-difference between *possessing* a reliable true belief and *possessing* a mere true belief obtains "even if there turns out to be no significant difference in value between the *belief components* of the states of affairs in question" (2007, 345, Olsson's emphasis). The first conditional probability solution, then, violates condition (5).

Olsson's second conditional probability solution, however, satisfies this condition. The second solution rests on Olsson's Reliability-Stability Thesis: "The probability that S's belief that  $p$  will stay in its place is greater, conditionally upon S's having a reliably acquired true belief that  $p$ , than it would be conditionally upon S's having a mere true belief that  $p$ " (2007, 347). This is Olsson's rendering of a principle stated by Williamson (2000, 79), which also expresses the durability of knowledge over mere true belief. Thus, reliable beliefs are supposed to have staying power, more stability. Setting aside the question why they're more stable, it's important here to note why stability is more *valuable*. Olsson argues that true beliefs with staying power have greater *practical utility* – they are more conducive to successful action – than mere true beliefs, which are less stable. If you want to go to Larissa, you are more likely to get there if you have a reliable (stable) true belief about which way to go. So, in accordance with condition (5), this solution identifies a value-difference between reliable true belief and mere true belief *at the level of the beliefs themselves*. However, the value-difference this account secures is a difference in *practical utility*, not a difference in *epistemic value*, as Olsson himself notes (2007, 347). This solution, then, violates condition (4) of the swamping problem.<sup>5</sup>

The final reliabilist proposal to consider is the value-autonomization solution due to Goldman and Olsson (2009). What is autonomous value? Whether or not a property (or type) has autonomous value is a function of the value-attributions (or value-attributitional tendencies) of a population of people at a time. Abstractly, if some property,  $P$ , is viewed as leading to (or tending toward) some fundamental end,  $F$ , then people often ascribe value to  $P$ -instances regardless of whether or not  $F$  is present. In at least some of these cases, people also ascribe higher value to  $P$ - $F$  complexes than they do to instances of  $F$  alone. If this happens,  $P$  has acquired autonomous value in that population. Goldman and Olsson offer the example of moral motives. How is it that some motives come to have moral value? Goldman and Olsson suggest that morally good motives derive their value from the good actions that they typically produce, while good actions have fundamental value (2009, 33). Obviously, such motives do not *always* lead to good actions. But according to Goldman and Olsson they do so often enough that the motives themselves come to *inherit* the value fundamentally attached to morally good actions. A motive like compassion that generally leads to good actions is still regarded as a *good motive* even on the occasions when it fails to produce a good action. More importantly, people see good deeds done out of good motives as morally

---

<sup>5</sup> Olsson (2007, 347) mentions the possibility of an account of the epistemic, as opposed to pragmatic, value of reliabilist knowledge based on stability, referring in an endnote to an unpublished paper by Erik Mohlin.

better than other good deeds. For instance, an act of compassion that alleviates suffering is generally regarded as better, morally speaking, than an act that unintentionally alleviates suffering. The motive inherits autonomous value from the good acts it typically produces, and common intuitive judgments ascribe higher value to the motive-act complex than to the act alone (or the same act done from a non-good motive).

The first problem with the value-autonomization solution is that it fails condition (6) of the swamping problem. That is, it doesn't explain how reliability enhances the epistemic value of true belief; rather, it sketches a psychological explanation of how people might come to attribute greater value to reliable true beliefs than to mere true beliefs. Consider why the problem is a deep one. For starters, not all cases of (acknowledged) instrumental value undergo autonomization. Goldman and Olsson (2009, 32) note, for instance, that taking aspirin is viewed as an effective means of headache relief; yet since aspirin-induced headache relief is not regarded as better than spontaneous headache relief, taking aspirin doesn't enjoy autonomous value. So which is reliability like – moral motives or taking aspirin? The question is an entirely empirical one, and depends on our value-attribitional tendencies. Assuming that reliability is instrumental in getting true beliefs, do people tend to assign reliable true beliefs higher value than mere true beliefs? Goldman and Olsson don't provide any evidence that bears on this question, however we know there is a significant population, reliabilism's critics, whose valuational practices do not ascribe autonomous value to reliability. This raises a number of troubling questions: If there are different groups, viz., those who do and those who do not ascribe autonomous value to reliability, how is this supposed to bear on the swamping problem? What are we to make of the different patterns of valuation in these groups? And even if "our" population ascribes higher value to reliable true beliefs than to mere true beliefs, isn't it a legitimate question whether we're correct to do so? Indeed, isn't this the question reliabilism's critics are pressing? If it is a legitimate question, then pointing out that reliability enjoys autonomous value for "us" – being merely a description of our valuational practices – simply fails to answer the critics' questions, and so fails to solve the swamping problem.

### 3. A reliabilist solution that satisfies the requirements.

The objective of this section is to explain how reliability enhances the epistemic value of a true belief. I will do so by developing a novel version of Olsson's conditional probability solution. To begin, recall the rationale behind ST – the thesis that reliable true belief has no more value than mere true belief. It claims that while reliability may have value, its value is derived entirely from the value of truth, so a belief that is both

reliably formed and true has no more value than a belief that is merely true. The assumption behind ST, then, is that if reliability is valuable *only* for its connection to truth, then the value of truth swamps that of reliability – i.e., there is no value-difference between reliable true belief and mere true belief. But this assumption is false. To show this, I will assume that reliability is valuable *only* for its connection to truth, and then explain why reliable true belief nevertheless has greater epistemic value than mere true belief. The thrust of the account here is that reliable true belief has a *closer* connection to the epistemic end than does mere true belief. This belies the claim, presented in section 1, that reliable true belief and mere true belief equally constitute the epistemic end of believing truly and avoiding error. This seems counterintuitive. How can it be possible for two true beliefs to satisfy the epistemic end to different degrees, that is, for one true belief to bear a closer connection than the other to the epistemic end of believing truly and avoiding error?

The account I develop here rests on an underappreciated feature of the epistemic end. Specifically, the aim to attain truth while avoiding error is not merely two-dimensional; that is, the aim of belief is not simply to attain truth and avoid error here and now. Rather, the epistemic aim has a diachronic character: It is to have and keep the truth on matters that concern us. One way to put this is to say that the aim is to believe truths in such a way that so long as a belief stays true it stays a belief, and if a belief loses its truth, then it is dropped. Putting the matter this way requires some attention to how beliefs are individuated over time. Saying that a belief might remain true over time, or at some point no longer be true, for instance, suggests that beliefs are not to be analyzed as dated propositions. Dated propositions, if true, are true eternally, and so there is no point at which they are no longer true. A complete account of the identity-conditions for beliefs over time is not in the offing here, but continuity of syntactic and semantic properties seems to be chief among them. Ordinary locutions like, 'Sam still believes in fairies,' do not attribute belief in the *continued existence* of fairies, but the *continued belief* that fairies exist. What makes Sam's belief last year and Sam's belief this year the same belief is (among other things perhaps) the continuity of the syntactic and semantic properties of these belief-episodes. On this view, Karen's true belief that her keys are on the table in the next room becomes false when unbeknownst to her the cat knocks them to the floor. Unless she checks into it, her belief will typically persist for a time. On the dated-propositions approach, Karen's belief at  $t_1$  (before the fall) and her belief at  $t_2$  (after the fall) are two different beliefs which vary in truth-value. But it is clear that there is a robust sense in which beliefs persist through time. Karen will continue to believe that her keys are on the table, even after they've fallen to the floor. Connected with the diachronic character of the epistemic end, this approach to the identity-conditions of beliefs helps shed light on how reliability



contributes value to a true belief, even as reliability derives its value only from its connection to truth.

We begin by returning to Olsson's second conditional-probability solution, which was based on the Reliability-Stability Thesis (RST): "The probability that  $S$ 's belief that  $p$  will stay in its place is greater, conditionally upon  $S$ 's having a reliably acquired true belief that  $p$ , than it would be conditionally upon  $S$ 's having a mere true belief that  $p$ " (2007, 347). The solution based on this thesis was rejected in section 2 because the value-difference between reliable and mere true belief it secured was pragmatic, not epistemic. But there's a more integral problem with RST. Namely, it works only if limited to beliefs about facts that are themselves quite stable. If someone has a reliable true belief today about which of two roads leads to Larissa, that belief will likely still be there, and still be true, tomorrow and the next day. But RST is false, or in any case does not point to a good-making feature of reliability, when applied to beliefs about quite changeable states of affairs. For instance, that the sun is at its zenith and that these berries are almost ripe are fleeting states of affairs. So it would not be a good-making feature of reliability (not even good for successful action) if a reliable true belief about a *fleeting* state of affairs were more *stable*, in the sense of being more likely to stay put, than a mere true belief. If such beliefs "stay put," they do not stay true (for long). The sun will no longer be at its zenith in fifteen minutes; the berries won't be "almost ripe" in a few days. Thus, RST is only adequate to beliefs about stable states of affairs, and so only gets things half-right.

The mistake of emphasizing the stability of reliable belief may be corrected by revising RST to an alternative conditional probability thesis:

Reliable Maintenance Thesis (RMT): The probability that  $S$ 's true belief will be appropriately maintained over time is greater conditional on its being reliably formed than on its being unreliably formed.

A belief's being appropriately maintained over time is to be determined relative to the epistemic end, now understood diachronically. If the epistemic end is to believe truths in such a way that so long as a belief stays true it stays put, and if a belief becomes false it is dropped or revised, then a belief is appropriately maintained over time insofar as it satisfies this end. In short, a belief is appropriately maintained over time if it is accepted when true, and revised when false. Stable beliefs are appropriately maintained only if they are beliefs about stable facts; in these cases, the beliefs stay put and stay true. Stable beliefs are not appropriately maintained if they are about fleeting states of affairs, because they stay put even when they are no longer true. According to RMT, if  $S$  has a reliable true belief one day, and the belief is no longer true the next, there is a

greater probability that S will drop the belief than there would be were S's belief a mere true belief to begin with. Whether RMT is correct will be addressed shortly. But, if true, RMT has two clear advantages over RST: (1) RMT is more general than RST; it entails RST when restricted to beliefs about stable states of affairs, but it is also adequate to beliefs about fleeting states of affairs. (2) RMT directly engages the swamping problem by drawing an explicit connection between reliability and the epistemic end; by contrast, RST draws a connection between reliability and stability, which Olsson then uses to further connect to the end of successful action, not the epistemic end.

The question, then, is whether RMT is correct. Like Olsson's case for RST (2007, 348-349), the case for RMT is an empirical one. Specifically, it is based on three empirical considerations. The first is that there is an utterly non-miraculous connection between those segments of the world about which a reliable process forms beliefs (its domain) and its success at forming true ones. Reliable processes are causally responsive to objects and properties in their respective domains. This causal connection determines a process's *propensity* to form true beliefs, and a process is reliable just in case its propensity to form true beliefs is sufficiently high. The second empirical consideration is that both reliable processes and their respective domains are fixed enough that this non-miraculous connection is very likely to hold into the future. If so, then processes that are reliable in the particular domain of one's current epistemic environment are likely to continue to be reliable, and unreliable ones are likely to remain unreliable. Olsson's defense of RST appeals to a similar principle, which he calls *generality*: "a method that is reliable in one situation is likely to be reliable in other similar situations in the future" (2007, 348). Finally, and crucially, RMT depends on the following statistical claim: if S uses reliable processes to arrive at the truth on some matter, then she is statistically more likely to use reliable processes to *maintain* and *update* beliefs in this domain in the future than she would be had she acquired the truth from an unreliable source. Therefore, if S uses reliable processes to form a true belief, B, then she is more likely to use reliable processes to *maintain* and *update* B than she would otherwise be. Why should this be true?

To simplify the account, assume that talk of 'maintaining' a belief always refers to its storage in short- or long-term memory, and also assume that short- and long-term memory are conditionally reliable, that is, true beliefs tend to be retrieved from memory when true beliefs are entered in memory. A consequence of this simplifying assumption is that both reliable true beliefs and mere true beliefs, if stored in memory, are reliably *maintained*. So why believe that reliable true beliefs are more likely to be reliably *updated* than mere true beliefs? In the commonest cases, if S uses a reliable process to form a belief, then she is statistically more likely to use *the very same process* in updating it. I take this to be a simple consequence of the functional economy of

human psychology. Consider, for instance, updating one's belief about the ripeness of some berries or the weekend's weather forecast. The process or method originally used to form the belief (*checking the berries' color, consulting a meteorologist's report*) is also the most likely to be used in updating it. Moreover, if S uses an unreliable process to form a *true* belief, she is most likely to redeploy that process in updating it. (It may be less likely to be reused if it leads to an egregious falsehood.<sup>6</sup>) Thus, that one is most likely to update true beliefs with the same processes used to form them in the first place supports the needed claim that if S uses a reliable process to form a true belief, she is statistically more likely to use reliable processes in updating it. The same cannot be said for beliefs formed by unreliable processes. If S uses unreliable processes to form a true belief, she is not more likely to use reliable processes in both maintaining and updating it. If a process was unreliable when she formed the belief, then it is likely to remain unreliable when she updates the belief. So even if one happens upon a truth on some matter, one is less likely to keep track of it. This establishes RMT, the thesis that the probability that S's belief, B, will be appropriately maintained (held if true, revised if false) over time is greater conditional on its being a reliable true belief than on its being a mere true belief.

If the account is correct so far, then the assumption behind ST, the claim that if reliability is valuable only for its connection to truth, then the value of truth swamps that of reliability, is false. This assumption gets its appeal by evaluating beliefs as one-shot attempts at the epistemic end (conceived of as a momentary state); evaluated in this way, both reliable true belief and mere true belief appear to *constitute* that end (in virtue of both being true beliefs). But I have shown that even granting that reliability is valuable only for its connection to the epistemic end of believing truly and avoiding error, reliability has a stronger connection to that end than mere true belief. This is so in virtue of (1) the diachronic character of the epistemic end and (2) the increased odds reliable beliefs have of being appropriately maintained over time. It is possible, then, for two true beliefs to satisfy the epistemic end to different degrees, and it is better, epistemically speaking, to have a reliable true belief than to have a mere true belief.

---

<sup>6</sup> This follows from a principle Olsson calls *learning*, according to which "a method that was unproblematically employed once will tend to be employed on similar problems in the future" (2007, 348). If problems are patent, then the process may be less likely to be redeployed in similar situations. I should add here that Olsson points out that his appeal to learning imposes an internalist requirement on his account of the (practical) value of reliability (2007, 351-352). My account of RMT does without internalist requirements. Olsson acknowledges that a conditional probability solution is generally compatible with externalism in his (2009, 108), but maintains (2009, 108n16) that the case for RST requires a "modest" internalism.

It may seem that this response merely pushes the swamping problem back a step, and that the problem can simply be re-cast in diachronic form. If the value of mere true belief swamps the value of reliable true belief when beliefs are evaluated as one-shot attempts at achieving the epistemic end, why doesn't the value of mere-true-belief-appropriately-maintained (MTBAM) swamp the value of reliable-true-belief-appropriately-maintained (RTBAM)? Surely it is *possible* for a mere true belief to be appropriately maintained (held if true, revised if false). Imagine two people, M and R, both of whom form some belief, B. M's belief is merely true, and R's is reliably true. Now assume that both maintain B appropriately. Isn't it the case that M's belief is just as good as R's belief vis-à-vis the epistemic end diachronically conceived? There are two ways to respond to this question. The first is simply to note that even if, when it comes to appropriately maintained beliefs, there is no value-difference between reliable and mere true beliefs, i.e.,  $V(\text{RTBAM}) = V(\text{MTBAM})$ , this wouldn't at all show that there is no value difference between reliable and mere true beliefs period, i.e.,  $V(\text{RTB}) = V(\text{MTB})$ . So even if the swamping problem re-emerges at the diachronic level (or for some other limited class of beliefs), the solution based on RMT remains intact.<sup>7</sup> But this all assumes that the swamping problem *does* re-emerge. Return to M's and R's two beliefs above. To think that the beliefs are epistemically equally good presupposes a *retrospective* notion of instrumental value, according to which a state's instrumental value is determined by its actual consequences, which can typically only be known after they occur. This is to be contrasted with a *prospective* concept of instrumental value, on which a state's instrumental value is determined by the probabilities of its consequences over a range of actual and possible cases. A state's instrumental value is not generally limited to the value of its actual consequences, but extends to its possible consequences – whether or not they materialize. For instance, we wouldn't say that a \$50,000 bank account and \$5,000 bank account have equal instrumental value if their owners, as a matter of contingent fact, spend the same amount out of them on the same things (even assuming they have the same needs and desires). Likewise, a properly functioning wrench and a broken wrench don't have equal instrumental value just because, as a matter of contingent fact, both of them have been left, unused and forgotten, in the back corner of a tool shed. The \$50,000 bank account and the properly functioning wrench *could* be used for things they're not being used for, and therein lies their greater instrumental value. A reliable true belief has a higher antecedent probability of being maintained appropriately across various actual and counterfactual scenarios than any mere true belief that is "accidentally" appropriately maintained over time. Thus, the reliable true belief has higher epistemic value (instrumental value vis-à-

---

<sup>7</sup> I thank an anonymous referee for *dialectica* for helping me see this point.

vis the epistemic end) than the mere true belief. In light of this, there is no chance of the swamping problem re-emerging in diachronic form. For a mere true belief to have epistemic value equal to a reliable true belief, its probability of being maintained appropriately over time and across various actual and counterfactual scenarios must equal that of a reliable true belief. But then it is no *mere* true belief.

#### 4. The swamping problem solved?

In the last section, I presented an account of GVRK that satisfies the six conditions of reliabilism's critics detailed at the end of section 1. At the heart of this solution is RMT, which states that the probability that *S*'s true belief will be appropriately maintained over time (vis-à-vis the epistemic end) is greater conditional on its being reliably formed than on its being merely true. According to condition (1), the solution grants that GVK, REL, and ST are genuinely logically inconsistent. According to condition (2), the solution compromises neither GVK nor REL. And according to condition (3), the solution explains how reliability enhances the value of true belief, that is, it explains why ST must be false. Since this explanation doesn't cite our value-attribitional tendencies, it accords with condition (6) as well. Moreover, in accordance with condition (4), the value reliability imparts to true belief on this solution is distinctively epistemic value – instrumental value toward the epistemic end. And in accordance with condition (5), this value accrues to the belief itself and not some more inclusive state of the epistemic agent who holds the belief.

Are reliabilism's critics likely to be satisfied that there is an adequate reliabilist solution to the swamping problem? Unfortunately not, and the reason is that the epistemic value this solution secures for reliabilist knowledge is contingent on certain empirical considerations. In a recent paper, Kvanvig claims that "the core of the swamping problem" is the view that "once the relevant controls are in place, we should expect, *always and necessarily*, for knowledge to be a value-enhancing characteristic of a state of true belief" (2010, 101, my emphasis). The controls Kvanvig is talking about are other, non-epistemic values that might override the value of knowledge. For instance, if knowing one more fact would bring about the demise of my friends and family, then it would be better, all things considered, for me to make due with a mere true belief instead. The solution elaborated in section 3 does not entail that reliable true belief always and necessarily has greater epistemic value than mere true belief. My solution rests on the assumption that once you've used a reliable process to form a true belief, you're more likely to use it in updating the belief. It rests on the assumption that processes that have been reliable in the past will continue to be so. These statements are not *necessarily* true. And they may not *always* be true. But the claim that knowledge

is always and necessarily more valuable than mere true belief (or indeed any other categorical imperative) is difficult to argue for, at best, and risks disengaging Kvanvig's naturalistically-minded opponents, thereby demoting the swamping problem from its status as a substantive and interesting challenge for reliabilism. I set out to give a reliabilist solution to the swamping problem that concedes as much as possible to reliabilism's critics. But perhaps reliabilism's critics need to make a few concessions as well. For instance, reliabilists, *qua* naturalistic philosophers, should not be expected to accommodate a world of objective, necessary values – a world of categorical imperatives. This in no way trivializes the swamping problem, for, as I have argued, none of the existing reliabilist solutions satisfy all the other conditions. The solution I've offered secures a naturalistically tractable kind of value for reliable true belief – instrumental value vis-à-vis the epistemic end. Moreover, it directly engages, and undermines, the most prevalent rationale for ST, which is the intuition that if reliability is valuable only for its connection to truth, then the value of truth swamps that of reliability. All the while respecting most, if not all, of the basic terms of the swamping problem required by reliabilism's critics. In conclusion, the swamping problem is effectively, and reliably, dammed.<sup>8</sup>

---

<sup>8</sup> I thank David Cole, Isaac Glahn, Peter Kung and anonymous referees for *dialectica* for their helpful comments on earlier drafts of this paper. I thank Hanover College for release time to work on this project.

## References

- Brogaard, B. 2006, "Can Virtue Reliabilism Explain the Value of Knowledge?", *Canadian Journal of Philosophy* 36, pp.335-354.
- Goldman, A. I. and E. J. Olsson. 2009, "Reliabilism and the Value of Knowledge" in D. Pritchard, A. Millar, and A. Haddock, eds., *Epistemic Value*, Oxford: Oxford University Press, pp.19-41.
- Greco, J. 2002, "Virtues in Epistemology" in P. K. Moser, ed., *The Oxford Handbook of Epistemology*, New York: Oxford University Press, pp.287-315.
- \_\_\_\_\_. 2003, "Knowledge as Credit for True Belief" in M. DePaul and L. Zagzebski, eds., *Intellectual Virtue: Perspectives from Ethics and Epistemology*, New York: Oxford University Press, pp.111-134..
- Kvanvig, J. L. 2003, *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press.
- \_\_\_\_\_. 2010, "The Swamping Problem Redux: Pith and Gist" in A. Haddock, A. Millar, and D. Pritchard, eds., *Social Epistemology*, Oxford: Oxford University Press, pp.89-112.
- Jones, W. 1997, "Why Do We Value Knowledge?", *American Philosophical Quarterly* 34, pp.423-440.
- McBrayer, J. P. 2007, "Process Reliabilism, Virtue Reliabilism, and the Value of Knowledge", *Southern Journal of Philosophy* XLV, pp.289-302.
- Olsson, E. J. 2007, "Reliabilism, Stability, and the Value of Knowledge", *American Philosophical Quarterly* 44, pp.343-355.
- \_\_\_\_\_. 2009, "In Defense of the Conditional Probability Solution", *Grazer Philosophische Studien* 19, pp.93-114.
- Riggs, W D. 2002, "Reliability and the Value of Knowledge", *Philosophy and Phenomenological Research* 64, pp.79-96.
- Sosa, E. 2007, *A Virtue Epistemology*, v.1, Oxford: Clarendon Press.
- Swinburne, R. 1998, *Providence and the Problem of Evil*, Oxford: Clarendon Press.
- Williamson, T. 2000, *Knowledge and its Limits*, Oxford: Oxford University Press.
- Zagzebski, L. 1996, *Virtues of the Mind*, Cambridge: Cambridge University Press.
- \_\_\_\_\_. 2000, "From Reliabilism to Virtue Epistemology" in G. Axtell, ed., *Knowledge, Belief and Character*, Lanham, MD: Rowman and Littlefield, pp.113-122.
- \_\_\_\_\_. 2003, "The Search for the Source of Epistemic Good", *Metaphilosophy* 34, pp.12-28.