Reliabilism, Stability, and the Value of Knowledge

Olsson, Erik J

Published in:
American Philosophical Quarterly

2007

Link to publication

Citation for published version (APA):
Olsson, E. J. (2007). Reliabilism, Stability, and the Value of Knowledge. *American Philosophical Quarterly*, *44*(4), 343-355.

Total number of authors:
1

Reliabilism, Stability, and the Value of Knowledge

Erik J. Olsson

Olsson, E. J. (2007). Reliabilism, Stability, and the Value of Knowledge. American Philosophical Quarterly, 44, 343-355.

Abstract: According to reliabilism, knowledge is basically true belief acquired through a reliable process. Many epistemologists have argued recently that reliabilism fails to accommodate our pre-systematic judgment that knowledge is more valuable than mere true belief. The paper pinpoints where this so-called swamping argument goes wrong. It is then argued that true beliefs that are reliably acquired are more stable and therefore more valuable for the purposes of guiding practical action over time. Finally, it is suggested that the stability thesis can, to some extent, bridge the gulf between externalist and internalist approaches to epistemology. While knowledge may be best defined in externalist terms, the full realization of its value requires the satisfaction of an internalist condition of track-keeping stating that people maintain a record of how their beliefs were acquired.

1. Introduction

Knowledge, as Plato was the first to point out, is more valuable than mere true belief.[1] Any account of knowledge that failed to make room for this common-sense observation would be defective. Recently, process reliabilism, or reliabilism for short, has been criticized precisely on these grounds. Reliabilism is the view that a subject S knows that p if and only if (1) p is true, (2) S believes p to be true, (3) S's belief that p was produced through a reliable process, and (4) a suitable anti-Gettier clause is satisfied.[2] In the following, the focus will be on what may be called *simple reliabilism* as captured by conditions (1) – (3). The anti-Gettier clause will play no role in this paper. According to the objection, reliabilist knowledge and mere true belief turn out to be equally valuable. Thus Ward Jones (1997) writes:

> In short, given the reliabilist's framework, there is no reason why we should care what the method was which brought about a true belief, as long as it is true. We value the better method, because we value truth, but that does not tell us why we value the true beliefs brought about by that method over true beliefs brought about by other less reliable ones (p. 426).

Richard Swinburne (1999) makes a similar point:

> Now clearly it is a good thing that our beliefs satisfy the reliabilist requirement, for the fact that they do means that … they will probably be true. But, if a given belief of mine is true, I cannot see that it is any more worth having for satisfying the reliabilist requirement. So long as the belief is true, the fact that the process which produced it

> usually produces true belief does not seem to make that belief any more worth having (p. 58).

Finally, Linda Zagzebski (2003) rejects reliabilism on the basis of the following analogy:

> [T]he reliability of the source of a belief cannot explain the difference in value between knowledge and true belief. One reason it cannot do so is that reliability per se has no value or disvalue … The good of the product makes the reliability of the source that produced it good, but the reliability of the source does not then give the product an additional boost of value … If the espresso tastes good, it makes no difference if it comes from an unreliable machine … If the belief is true, it makes no difference if it comes from an unreliable belief-producing source (p. 13)

Similar objections have been raised by Wayne Riggs (2002), Jonathan L. Kvanvig (2003), and Ernest Sosa (2003). The main idea behind these criticisms is that while reliability is valuable because reliably acquired beliefs are mostly true, it does not add value once the belief produced by the reliable process is true. Once a belief is true, it doesn't become more valuable, "more true" if you will, as the effect of having been reliably produced. Some authors (e.g. Kvanvig, 2003) express this by saying that the value of reliability is "swamped" by the value of truth. Accordingly, the argument put forward by Jones, Swinburne et al is sometimes referred to as the *swamping argument*.[3]

The swamping argument is not merely an argument to the effect that reliabilist knowledge is no more valuable for trivial truths such as “There are *n* grains of sand on the beach” where *n* is the actual number of grains. That would be unsurprising. Rather, the swamping effect is assumed to set in also for propositions that matter to us. Even for such important propositions, reliabilist knowledge is no more valuable than mere true belief, or so the swamping theorist claims.

The swamping argument is not an argument against reliabilism *per se* but targets its combination with *veritism*, the view that true belief, and true belief only, has final or intrinsic epistemic value. It is not enough for the argument’s sake that “we value truth”, to use Jones’s liberal formulation. Our valuing truth is compatible with our valuing other things epistemic as well, like reliable production. If we assign reliable production final epistemic value, the swamping argument clearly doesn’t work, for the sum of the values of truth and of reliable production will then exceed the sum of the values of truth and of unreliable production. Rather, it is essential that reliability has no value in itself. A distinguished advocate of a reliabilist-veritist theory, Alvin I. Goldman has been the primary target of the swamping theorists’ efforts.[4]

As it stands, veritism is not without its problems. If getting at the truth is the only thing we value, then we don’t value avoiding falsehood. But we do seem to value avoiding falsehood. It is better not to believe p than to believe p, if p is a false proposition. Indeed, believing falsehoods presumably has negative value. A more plausible form of veritism would have to accommodate these observations. Yet if the swamping argument goes through for veritism in its original form, then it does so also for more subtle versions that attribute value to falsehood avoidance. After all, the swamping argument does not involve any reference to false beliefs. We are simply asked to focus on a *true* proposition p and to compare the value of knowing that p with the value of merely believing that p.

The problem to which the swamping theorist calls attention is more general than it might seem to be on first sight. Similar swamping arguments can be raised against competing accounts of knowledge, such as internalism. Consider an internalist theory according to which having justification has no other value than to indicate the truth of the belief thus justified. Paraphrasing Swinburne, the following objection could be leveled against the combination of such an internalist theory with veritism: “Now clearly it is a good thing that our beliefs satisfy the justification requirement, for the fact that they do means that they will probably be true. But, if a given belief of mine is true, I cannot see that it is any more worth having for satisfying the justificationalist requirement. So long as the belief is true, the fact that beliefs that are justified are usually true does not seem to make that belief any more worth having."

There is a broad consensus in the literature that the swamping argument is a knockdown argument against reliabilism which has thereby been shown to be clearly untenable. Some of these authors (e.g., Swinburne and Kvanvig) find reasons to prefer some form of internalism. Certain versions of internalism can, they maintain, account for the greater value of knowledge over true belief. And some (e.g. Kvanvig, Sosa, Riggs and Zagzebski) believe that virtue epistemology holds special promise for solving the value problem. According to virtue epistemology, in its basic form, S knows that p only if S acquired her belief in p by exercising some epistemic virtue, so that a person who knows can be credited for his or her true belief in a way in which a person who has a mere true belief cannot. The purpose of this paper is to show that the value problem *per se* is no good reason to give up either reliabilism or veritism.

2. Why the belief-espresso analogy fails

Let us return to Zagzebski and her belief-espresso analogy because it advances the swamping view in a particularly transparent manner**.** (The relevant passage from Zagzebski was quoted above.) Convincing as the analogy may seem, it fails in several respects to show that reliabilist knowledge lacks extra value in relation to mere true belief.

Here is the first point. In the cited passage, Zagzebski focuses on the final value of the *belief produced* that is here seen as an object comparable to a cup of espresso. Just as reliable production of a good espresso doesn't add to its hedonistic value, so too (assuming veritism) the epistemic value of a true belief is not enhanced by the fact that the belief was reliably produced. Learning that the thing (espresso or belief) was reliably produced does not make us value it more if we knew at the outset that the thing had final value. Let us grant this for a moment. It is still true, though, that learning that the thing was reliably produced may come as a pleasant surprise, for we now know that we are in a more fortunate *overall position* than we had reason to believe before. The machine or method that produced a given thing can usually be reemployed, and, if it is reliable, it will in all likelihood produce more things of final value. The likelihood of successful reemployment is lower if the machine or method is unreliable.

We can express the observation just made in terms of the relative value of *states of affairs*, namely, by saying that believing something truly as the result of a reliable process may be more valuable than believing something truly as the result of an unreliable one. Believing something truly on the basis of a reliable process may be more valuable in the sense that believing truly on that basis makes the obtaining of further states of true belief, things of final value, more likely. This could be so even if turns out to be no significant difference in value between the *belief components* of the states of affairs in question. The author has argued this point elsewhere and will not go into the details here.[5]

What this suggests is that Zagzebski is relying on an implicit premise, that the problem of the value of knowledge concerns exclusively the value of the product belief, seen as an object comparable to a cup of espresso, and not the value of knowing and merely truly believing seen as states of affairs. Prima facie, however, neither way of looking at value seems clearly more fundamental or more correct than the other. This shows that Zagzebski's analogy argument, however compelling *prima facie,* certainly falls short of being a *knockdown* argument.[6]

Now it is true that a good espresso doesn't taste any better in virtue of having been reliably produced. Similarly, if a belief is already true, it doesn't become "more true" in virtue of having been reliably produced. Learning that a belief was reliably produced doesn't make us more confident, if our degree of confidence was already at its maximum. The second point is that there are, nonetheless, reasons to think that a true belief becomes more *stable* for having been reliably produced.

Suppose you use an *un*reliable method to arrive at the belief that p, where p happens to be a true proposition. Normally you will have the same method at your disposal the next time the same kind of problem arises. So you will use the same method again. This time, however, the method, if it is unreliable, is relatively likely to produce a false belief (more so than if it had been reliable). If the new belief is indeed false, its falsity will normally be detected in the fullness of time. When this happens, you are likely to question other beliefs arrived at through the same method, including your belief that p.

Suppose (to take a modern version of Plato’s example) that you are traveling by car to Larissa and relying on the on-board navigation computer for geographic guidance. The navigation system, we assume, is unreliable but happens to give a correct recommendation at the first junction, say, that Larissa is to the right. Since the system is unreliable, it is likely eventually to give an incorrect recommendation; at least this is more likely than if it had been reliable. Moreover, the incorrectness of a recommendation is something that you are likely to detect. If, for instance, the road suddenly ends in the middle of nowhere, you will conclude that the system gave an incorrect recommendation. If it did, it is to some extent unreliable. If it is unreliable, previous recommendations may be wrong. In particular, the recommendation at the first junction may be wrong, and so you may need to retract your (true) belief that Larissa is to the right. This reply to the swamping argument will be examined in greater detail in the next section.

Nothing of the sort seems true for espressos. Suppose that the unreliable espresso machine that happens to produce a fine espresso on the first occasion produces, upon reemployment, an espresso that is barely drinkable. The existence of the second bad espresso does not in any intelligible sense destabilize the first (good) espresso in a way analogous to how the existence of the second false belief destabilizes the first (true) belief. The second false belief makes the first (true) belief disappear. On the espresso analogy, the second bad espresso should make the first (good) espresso disappear, but it doesn't, so the espresso analogy is false.

Against this the following objection could be raised: Isn't the difference here that you can know that the first espresso was good (by tasting it); whereas you did not really know that the first belief (which happened to be true) was indeed true? To the extent that the truth of the first belief (after employing the unreliable method to arrive at it) might be subsequently borne out or confirmed by *other* methods, it would not be destabilized by the discovery that the second belief was false, so the two cases are parallel after all.

The source of the dispute is that Zagzebski's analogy admits of several different interpretations. On the alternative reading upon which the objection is based the relevant contrast is between (a) the reliable production of a true belief *whose truth can be independently confirmed* and (b) the reliable production of a good espresso *whose goodness can be independently confirmed* (by tasting). The difficulty with this reading is that it makes the analogy irrelevant, or at least not directly relevant, to the claim Zagzebski wants to underpin. After all, the analogy is intended to substantiate the general claim that "the reliability of the source of a belief cannot explain the difference in value between knowledge and true belief" (op. cit.). But what is being compared on the current rendering is reliable production of *independently confirmed* true belief vs. reliable production of *independently confirmed* good espresso. There is, once more, no clear relevance here to Zagzebski's general claim. For the record, there is a further construal according to which reliable production of true belief is compared with reliable production of good espresso whose goodness can be independently confirmed (again, by tasting). Making the espresso example disanalogous from the start, this reading can be quickly dismissed as too uncharitable. This leaves us with the most reasonable, and perhaps most straightforward, interpretation: what is being compared is simply reliable production of true belief vs. reliable production of good espresso, and nothing is being assumed about the possibility of independent confirmation. For all that is known there may or may not be someone there to taste the espresso, and for all that is known there may or may not be someone there to verify the belief. The point that was made a few paragraphs ago was that on this understanding, there is still a problem having to do with stability.

Veritism may well be able to accommodate the greater epistemic value of true beliefs that persist over true beliefs that do not.[7] If this is not an option, the reliabilist-veritist can still argue that true beliefs that are stable are more valuable for purely practical reasons. Veritism is a thesis about epistemic value and is, as such, compatible with just about any view on what has practical or non-epistemic value. As vigorously argued by Timothy Williamson (2000), having stable true beliefs promotes successful action over time, if the success of your action depends on the belief in question being true. If you have a stable true belief as to where Larissa is, one that persists throughout your journey, you are more likely to get there than if your true belief is retracted somewhere along the way.[8]

In the next section, a closer look is taken at the stability thesis and the cognitive and other conditions that need to be satisfied in order for reliabilist knowledge to attain its distinctive practical value.

3. Reliabilist knowledge as promoting successful action over time

On the view sketched so far, reliablist knowledge, in addition to being epistemically more valuable than mere true belief, also has a distinctive practical value. Reliabilist knowledge, it is maintained, is conducive to successful acting over time in the sense that the probability that S will successfully complete an action over time whose success depends on p being true is higher, conditionally upon S's having reliabilist knowledge that p, than it would be conditionally upon S's having a mere true belief that p. For example, the probability that S will find her way to Larissa is greater, conditionally upon S's having reliabilist knowledge as to where Larissa is, than it would be conditionally upon S's having a mere true belief to the same effect. Using the standard notation for conditional probability:

P(S will get to Larissa | S has a reliably acquired belief as to where Larissa is)
>   P(S will get to Larissa | S has a mere true belief as to where Larissa is).

The argument for this claim, which we may call the Reliability-Action-Thesis (RAT), has two parts. The first part involves showing that reliabilist knowledge is conducive to stability of belief. 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, i.e.,

P(S’s belief that p will stay put | S believes truly that p due to a reliable process)
> P(S’s belief that p will stay put | S believes truly that p due to an unreliable process).

This is the Reliability-Stability-Thesis (RST). According to the second part of the argument, stability promotes successful acting over time. For instance,

P(S will get to Larissa | S’s true belief as to where Larissa is stays put)
> P(S will get to Larissa | S’s belief as to where Larissa is will be lost).

This is the Stability-Action-Thesis (SAT). Together, RST and SAT imply RAT.[9]

SAT is a pretty trivial claim, so comparatively little effort will be spent on its defense. Suppose you embark on a journey to Larissa with a correct picture of where Larissa is. Clearly you will be more likely to reach Larissa if your geographical belief stays in place throughout your journey. You would be worse off if you lost your belief somewhere along the way, or if your belief turned into belief in the negation. In the first case, you would enter a state of confusion or, as the old American pragmatists used to say, “doubt”. And, obviously, matters would be even worse if your belief were replaced by a false belief as to Larissa’s location. Clearly, then, having a true belief of the relevant kind that stays put is something that is of great advantage when acting over time. The probability that you will reach Larissa (and in time) is raised by assuming that you have a true belief as to its location when you embark on the journey. That probability is further increased by assuming that your true belief persists. These considerations are sufficient to establish SAT beyond reasonable doubt.

The defense of RST is considerably more subtle. The claim to be justified is that reliable acquisition of true belief is conducive to stability. The main part of the justification amounts to showing that, if one is using an unreliable method to acquire a given belief, the unreliability will tend to be detected in due course. Once the method has proven to be unreliable, beliefs that were acquired by means of that method will tend to be discarded. By contrast, the chance that doubt will be shed on an actually reliable process is lower, and it is correspondingly less likely that beliefs arrived at by means of such a method will later be found questionable.

In order to make this likely, appeal will be made to some empirical background assumptions. In particular, it will be assumed that, while our inquirers may sometimes succumb to wishful thinking and other less reliable paths to belief, most of their belief-acquisition processes are in fact reliable. This will be expressed by saying that they are *overall reliable*. Furthermore, inquirers will be supposed to be *track-keepers* in the sense that they keep a record of the sources of their beliefs. According to a further assumption, inquirers view their beliefs as *corrigible*, meaning that they typically do not stick to their beliefs no matter what. More precisely, an inquirer who finds a given belief false is likely to question the reliability of the method by means of which that belief was formed. Moreover, once a given belief-acquisition method is classified by the inquirer as dubious, all beliefs that were obtained solely or mainly through the use of that method are also, to some extent, in doubt. These three conditions give expression to one sense in which the inquirers' cognitive faculties are "in good order", to use Timothy Williamson's phrase (2000, p. 79).[10]

Finally, we will also assume that the following conditions hold:

- *Non-uniqueness*: once you encounter a problem of a certain type, you are likely to face other problems of the same type in the future

- *Cross-temporal access*: a method that was used once is often available when similar problems arise in the future
- *Learning*: a method that was unproblematically employed once will tend to be employed again on similar problems in the future[11]
- *Generality*: a method that is reliable in one situation is likely to be reliable in other similar situations in the future

For instance, non-uniqueness is satisfied in the coffee scenario because most people want to have coffee more than once in their lifetime. Similarly, the problem of finding one's way arises time and again, or else people wouldn't buy expensive general-purpose navigation equipment for their cars. These empirical conditions are plausibly satisfied for methods in general, whether or not they concern the production of espresso, or of belief, or of something else.[12]

Why, then, should there be a tendency for unreliably acquired (true) beliefs to be discarded? Suppose S has acquired the true belief that p by means of an unreliable method—call it M. By non-uniqueness, S is likely to confront the same type of problem again. By cross-temporal access, S is likely to have access to M when this happens. By learning, S is likely to make use of M on this future occasion. However, since M is unreliable, it is relatively unlikely that M will produce a true belief the second time around. An unreliable navigation system may produce a correct recommendation once, but there is no guarantee that it will do so upon reemployment. (If, by contrast, M had been reliable on the first occasion then, by generality, it would probably have been reliable on the second occasion as well, producing a new true belief.) If the new belief is actually false, this will tend to be discovered by the inquirer's other, mostly reliable belief-fixation processes.[13] More carefully put: a clash is likely to arise between the belief produced by M and the beliefs produced by some reliable belief-forming process at S's disposal. Subsequent verification by some basic reliable process, such as vision at close range, will tend to settle the issue in favor of what the reliable process was reporting. Finding the new belief false, S will tend to question the reliability of M, the process whereby it was adopted – this follows from the corrigibility assumption. By track-keeping, S will note that her belief in p was also produced by means of M. By corrigibility again, S will now find p to be a proposition whose truth cannot be taken for granted any longer.

This concludes the defense of the thesis that reliabilist knowledge promotes successful action over time, a thesis that was seen to rely on two other claims: that reliabilist knowledge promotes stability and that stability is conducive to successful action over time. The second part was easily made plausible. The first part required a more elaborate defense. The main thrust of that defense was that, for inquirers whose cognitive faculties are working properly operating in circumstances characterized by non-uniqueness, cross-temporal access, and so on, a reliably acquired true belief is more likely to be retained than an unreliably acquired true belief—mainly because the unreliable method is relatively likely in due course to produce false belief, and the falsity of the belief is something that the inquirer is likely eventually to detect. Once the unreliability has been detected, other beliefs that were acquired by means of the same process will tend to become discredited as well.

4. Kvanvig on stability

Not everyone agrees that the added value of knowledge has anything to do with the stability of belief. In the very first chapter of his thought-provoking 2003 book, Jonathan L. Kvanvig resolutely dismisses the notion that beliefs that are known are more stable than beliefs that are merely true. The explicit target of his critique is Williamson's thesis to that effect. By 'knowledge' Williamson means a primitive state which cannot be defined in terms of other concepts. Thus, he rejects all attempts to provide an analysis of knowledge in terms of sufficient and necessary conditions, including the reliabilist construal of knowledge as, basically, reliably acquired true belief. Kvanvig is accordingly arguing against a view distinct from that advanced here. Nonetheless, the upshot of his discussion is that stability cannot explain the added value of knowledge *no matter how the latter concept is reasonably conceived*. As a consequence, when reliabilism and its swamping problem is discussed later on in Kvanvig's book (chapter 3), the stability reply is not taken up for discussion.

The purpose of this section is not to assess Kvanvig's criticism of Williamson but rather to evaluate his implicit assumption that his critique is general enough to show, by implication, that a reliabilist cannot avoid the swamping problem by appealing to the greater stability of true beliefs reliably acquired.[14]

Kvanvig's first point is that "knowledge, no less than true belief, can be lost" (p. 13). Translated into our framework, the claim is that reliably acquired true belief, no less than unreliably acquired true belief, can be lost. This is true, of course, but unproblematic. The thesis defended here is a comparative one which does not commit an advocate to the absolute stability of reliably acquired true belief. All it entails is the greater stability of that which is reliably known as compared to that which is merely believed truly in the conditional-probability sense explained above.

Kvanvig's second complaint is that the relevant comparative claim is "undermined if the true beliefs are thoroughly dogmatic ones" (p. 14). Does this critique carry over to the reliabilist case? Are mere true beliefs that are dogmatically held more stable than true beliefs that have been reliably acquired? Maybe they are. Still, the stability thesis advocated here is compatible with there being occasional merely true beliefs that are dogmatically held and therefore no less stable, or even more stable, than reliably acquired ones.

More precisely, the following claim may well be true:

(K) P(S's belief that p will stay put | S's true belief that p is dogmatically held) > P(S's belief that p will stay put | S's true belief that p was reliably acquired)

But K is perfectly consistent with the stability claim defended in this paper, viz.,

(RST) P(S's belief that p will stay put | S's true belief that p was reliably acquired) > P(S's belief that p will stay put | S's true belief that p was unreliably acquired)

The latter statement is that the proportion of stable beliefs among those that are true and reliably produced is greater than the proportion of stable beliefs among those that are true but unreliably produced. Surely, this statement about proportions can be true even though in some cases unreliably acquired true beliefs are as stable as, or even more stable than, reliably acquired ones. Just as noting that the claim that some non-birds (aeroplanes, for instance) fly has little bearing on the claim that birds are more likely than non-birds to fly, so too observing that some unreliably acquired true beliefs are stable has little bearing on the claim that reliably acquired true beliefs are more likely to persist than unreliably acquired ones. Kvanvig's subsequent reference to "beliefs fixed by mechanisms having survival value" (p. 15) are, for similar reasons, of little relevance to the issue at hand.

Finally, Kvanvig complains that, even if it is true that beliefs that are known are more stable than beliefs that are merely true, this will be a fact that is "highly contingent" (p. 17). Kvanvig seems to imply that to the extent that knowledge is more valuable than mere true beliefs, it should have this added value necessarily, i.e., in all possible worlds, and not just in some worlds, such as the actual world or nearby worlds. By contrast, we have seen that whether or not reliable acquisition contributes to stability depends on the empirical circumstances, e.g., on whether people are mostly reliable in what they believe, whether they tend to regard their beliefs as corrigible, and so on. We can, of course, imagine worlds where these conditions are not satisfied in sufficient degree, e.g., where most belief-fixation methods are unreliable, et cetera. It is true, then, that the stability thesis put forward in this paper is not necessarily true. But why, it may be asked, does the extra value of knowledge have to be something that pertains to knowledge necessarily? Why could it not be a contingent, or even “highly contingent”, matter?[15] There are no clear answers to be found in Kvanvig’s book.

Be that as it may, the main objective of this paper has been to show why the swamping argument fails. If the swamping argument were sound, it would show that reliabilist knowledge is *necessarily* no more valuable than mere true belief. The argument, after all, does not appeal to any empirical considerations that need to obtain but is presented as a piece of philosophical arm-chair reasoning. If it were to succeed, it would do so regardless of empirical circumstances, regardless of what possible world we are focusing on. Thus in order to show that the swamping argument is wrong it is sufficient to show that reliabilist knowledge *can* be distinctively valuable, that there *are* circumstances or possible worlds where reliabilist knowledge is more valuable than mere true belief. Surely, this has been accomplished.

What about our own world? The empirical conditions of overall reliability, non-uniqueness et cetera could be said, with some qualifications soon to be given, to describe our actual world, so that our world is, at least in a restricted sense, one where the distinctive value of reliabilist knowledge is realized, although strictly speaking focusing on such realistic conditions was not necessary given the limited aims of this study. The most controversial condition, from the standpoint of realism, is that of track-keeping, which has been debated by psychologists, computer scientists and philosophers alike. Psychological experiments have shown that human beings only rarely remember the reasons for their beliefs, and that they often retain beliefs even when their original evidential basis is completely destroyed.[16] Harman (1986), pp. 29-42, interprets these experiments as indicating that people do not generally keep track of the reasons for their beliefs, so that they cannot tell when new evidence undermines the basis on which some belief was adopted. The common-sense position seems to be that we do keep track of our evidence if it is important to do so and, in particular, if it is likely that we will later be held accountable for our view. None of the experimental findings cited by Harman or anyone else seems to indicate that the common-sense position should be wrong. Rather, the experiments are constructed in such a way that the subjects have no intrinsic interest in the beliefs themselves or in their defense. Since accountability is a major concern in science and politics, we should expect scientists and politicians to keep track of the evidential basis of their professional beliefs. At the very least, then, reliabilist knowledge has added value in our world in these and similar contexts.[17]

5. Implications for the externalist-internalist debate

Contrary to what most commentators have thought, the swamping argument is not one that needs to upset the reliabilist-veritist. The argument presupposes that what is at stake is exclusively the value of the belief itself, whereas it is at least as plausible to think that we should focus on the value of states of affairs. If we do, reliabilist knowledge emerges as being indeed more valuable than mere true belief. Even if we look at the value of the belief itself, on the model of a cup of espresso, the swamping conclusion is not forthcoming. A true belief becomes more stable as the effect of reliable production. Stability among true beliefs is practically valuable and may, from the point of view of a refined veritist position, also have a distinctive epistemic worth.

A final suggestion will be offered as to how the foregoing discussion may bear on the notorious externalist-internalist debate in the theory of knowledge. According to externalism, knowledge requires reliability or some other condition whose satisfaction need not be accessible to the subject. Thus from an externalist perspective one can know without also knowing that one does, e.g., without knowing that one's belief was reliably produced. The internalist, on the other hand, maintains that knowledge requires some sort of mental representation of the evidence or mechanisms upon which the belief is based. Arguments can be cited in favor of either view. Thus externalism, unlike internalism, is plausible as an account of observational knowledge and also of knowledge in animals and smaller children, whereas characteristically human (adult) knowledge, it is often argued, requires the satisfaction of an internalist condition.

The stability thesis defended in this paper states that a true belief becomes more stable as the effect of being reliably acquired. This however is not so in every conceivable situation. The thesis presupposes the holding of some identifiable empirical conditions. One of these conditions is track-keeping, stating that the person maintains a record of how a given belief was arrived at, i.e., of the type of belief-acquisition process that terminated in the belief in question. Only then can the subsequent discovery of the unreliability of a given fixation method lead to the discrediting of other beliefs previously fixed using that same method or process. Without track-keeping this is hardly possible.

Now these considerations are relevant here because *track-keeping is a modest internalist requirement on a cognitive agent*. It requires that the agent maintain a mental record, a record in her mind, of how beliefs were acquired. In that sense, track-keeping is an internalist requirement. It is a modest requirement because track-keeping is, to be sure, possible without the agent maintaining a record of the required sort *in her head*, as opposed to, say, writing it down on paper or storing it in a computer file. What is required is merely that the agent keep track of her beliefs in what has been called her "extended mind" (Payne, 1992, pp. 104-109; Norman, 1991), i.e., in a storage medium accessible to her.[18]

The requirement of track-keeping goes beyond the content of the externalist position. A person may have externalist-reliabilist knowledge without recording the origins of her beliefs. While track-keeping is not required by reliabilism per se, it is part of the cognitive environment in which reliabilist knowledge promotes stability of belief and thereby attains its full practical value. Hence, *even if knowledge is best defined in an externalist manner, the full realization of its value requires the satisfaction of a modest internalist condition*.[19]

This way of looking at the externalist-internalist debate has the merit of giving some credit to both camps in a way that makes their approaches look complementary. Externalists should be recognized for having produced a plausible *analysis* of knowledge, though without paying much attention to the conditions under which knowledge achieved its distinctive value. Internalists, one the other hand, while rightly emphasizing the importance of internal factors, have been mistaken about their exact role. Again, such factors, rather than entering into the conditions defining knowledge, are better seen as essential elements of the environment in which knowledge attains its maximum worth.[20]

References

Alchourrón, C., Gärdenfors, P., and Makinson, D. (1985), "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision", *Journal of Symbolic Logic* 50: 510-30.

Armstrong, D. M. (1973), *Belief, Truth and Knowledge*, Cambridge University Press.

Doyle, J. (1992), "Reason maintenance and belief revision: foundations vs. coherence theories", pp. 29-51 in *Belief Revision*, Gärdenfors, P. (ed.), Cambridge Tracts in Theoretical Computer Science 29, Cambridge University Press.

Goldman, A. I. (1976), "Discrimination and Perceptual Knowledge", *The Journal of Philosophy*, 73: 771-91.

Goldman, A. I. (1986), *Epistemology and Cognition*, Harvard University Press.

Goldman, A. I. (2002), *Pathways to Knowledge*, Oxford: Oxford University Press.

Goldman, A. I., and Olsson, E. J., "Reliabilism and the Value of Knowledge", forthcoming in Haddock, A, Millar, A, and Pritchard, D. (eds.), *Epistemic Value*, Oxford University Press.

Harman, G. (1986), *Change in View: Principle of Reasoning*, Cambridge, Mass.: MIT Press.

Hilpinen, R. (1995), "Belief Systems as Artifacts", *The Monist*, 78 (2): 136-155.

Jones, W. E. (1997), "Why do we value knowledge?", *American Philosophical Quarterly* 34, No. 4, October: 423-439.

Kvanvig, J. L. (2003), *The Value of Knowledge and the Pursuit of Understanding*, Cambridge University Press.

Levi, I. (1980), *The Enterprise of Knowledge*, Cambridge, Mass.: MIT Press.

Levi, I. (2004), *Mild Contraction: Evaluating Loss of Information due to Loss of Belief*, Oxford University Press.

Mohlin, E. (2006), "Veritistic Unitarianism and the Value of Stable Belief", unpublished manuscript.

Norman, D. A. (1991), "Cognitive Artifacts", in *Designing Interaction: Psychology of the Human-Computer Interface*, Caroll, J. (ed.), Cambridge University Press, pp. 17-38.

Payne, S. J. (1992), "On Mental Models and Cognitive Artifacts", in *Models in the Mind: Theory, Perspectives and Application*, Rogers, Y. (ed.), London: Academic Press, pp. 103-18.

Riggs, W. D. (2002), "Reliability and the Value of Knowledge", *Philosophy and Phenomenological Research*: 79-96.

Ross, L., and Anderson, C. A. (1982), "Shortcomings in the Attribution Process: On the Origins and Maintenance of Erroneous Social Assessments", in *Judgment under Uncertainty: Heuristics and Biases*, Kahneman, D. et al (eds.), Cambridge University Press, pp. 129-152.

Shogenji, T. (2003), "A condition for transitivity of probabilistic support", *British Journal for the Philosophy of Science* 54: 613-616.

Sosa, E. (2003), "The Place of Truth in Epistemology", in *Intellectual Virtue: Perspectives from Ethics and Epistemology*, Oxford University Press: 155-179.

Swinburne, R. (1999), *Providence and the Problem of Evil*, Oxford University Press.

Tennant, N. (2003), "Theory-contraction is NP-complete", *Logic Journal of the IGPL* 11 (6): 675-693.

Williamson, T. (2000), *Knowledge and Its Limits*, Oxford University Press.

Zagzebski, L. (2003), “The Search for the Source of Epistemic Good”, *Metaphilosophy*, 12-28.

[1] See Plato’s dialogue *Meno*.

[2] See Goldman (1976) and (1986) for classical formulations of the reliabilist view.

[3] It could be objected to the swamping argument that few reliabilist have claimed that knowledge is reliably acquired true belief period. Most, if not all, reliabilist also believe that there is a need for a fourth anti-Gettier condition. This opens up for the possibility that it is this fourth condition that is responsible for the greater value of knowledge over mere true belief. Still, the basic idea of reliabilism is that of knowledge depending on the existence of a reliable process, and it would be seriously damaging for the theory if that very feature failed to add value.

[4] As for Goldman’s veritism, see Goldman (2002), p. 53.

[5] For a detailed account of this response to the swamping argument, see Goldman and Olsson (forthcoming). A similar suggestion is made in passing by Armstrong (1973) in a different context (in response to an objection raised by Deutscher concerning the so-called generality problem for reliabilism). See also Williamson (2000) for a related proposal for a particular class of (temporally related) beliefs. Williamson, however, rejects all attempts to analyze knowledge, including reliabilism. Neither Armstrong nor Williamson addresses explicitly what has become known as the swamping problem.

[6] Percival (2003, p. 38) notes that the exact content of our pre-systematic judgment that knowledge is more valuable than mere true belief is “obscure”. He goes on to say that, “[f]or all we know at the outset, it amounts to no more than the claim that, by and large, for all rational agents x and propositions p, x prefers his knowing that p to his merely believing that p truly” (ibid.).

[7] Erik Mohlin (2006) proposes a way to measure veritistic value that takes into account the stability of a true belief.

[8] Cf. Williamson (2000), p. 7-8.

[9] The relation of support is not transitive in general; it is not generally true that, if X supports Y and Y supports Z, then X supports Z. Let X be “S is an academic philosopher”, Y be “S has a doctoral degree” and Z be “S is well paid”. Then X supports Y and Y supports Z but X fails to support Z. However, there are conditions under which transitivity in fact holds. As noted in Shogenji (2003), this happens when the intermediate proposition screens off the original evidence with respect to the hypothesis in question in the following precise sense: (1) $P(Z|X \& Y) = P(Z|Y)$ and (2) $P(Z|X\&\neg Y) = P(Z|\neg Y)$. To see that these conditions are plausibly satisfied in our case, let X be “S’s true belief was reliably obtained”, Y be “S’s true belief is stable” and Z be “S’s will act successfully over time”. Clearly, once we know the truth value of Y, learning in addition that X is true does not affect our confidence in Z.

[10] In his elucidation of what it means for one’s faculties to be in good order Williamson stresses the importance of not entertaining “profoundly dogmatic beliefs” (2000, p. 79) corresponding to the assumption of corrigibility. The author is not aware of any discussion of Williamson’s on the role of overall reliability and track-keeping.

[11] As I use the expression, a belief generating method was “unproblematically employed” if there is no positive sign suggesting to the believer that the belief thus produced might be false.

[12] These conditions are discussed at greater length in Goldman and Olsson (forthcoming).

[13] Here one could add that the inquirer must also be in some degree curious and receptive of new evidence for the unreliability not to go unnoticed. If, for instance, the navigation system recommends the false road to Larissa, this will be detected only if the inquirer receives incoming perceptual evidence suggesting, say, that the road ends with no city in sight.

[14] In the following, those of Kvanvig's objections that rely on Gettier-style cases are disregarded. This concerns, above all, the mathematics and tsunami examples on pp. 15-17. As explained in the beginning, this paper is solely concerned with simple reliabilism, i.e., reliabilism without an anti-Gettier clause.

[15] Cf. Ward Jones (1997) who maintains that any added value "will come from contingent characteristics of knowledge" (p. 433).

[16] See Ross and Anderson (1982), pp. 147-149, for an overview.

[17] See Doyle (1992) for a defense of track-keeping from the perspective of computer science.

[18] Cf. Hilpinen (1995), p. 144: "The address book which contains information about the addresses and telephone numbers of my acquaintances can be regarded as part of my 'belief system,' and it seems epistemically irrelevant whether I carry it in my head or in my pocket."

[19] Some researchers maintain that track-keeping is a *sine qua non* for rational belief revision. See, for instance, Doyle (1992) and, more recently, the account of "dependency networks" in Tennant (2003). There is a diverging tradition in belief revision according to which keeping track of one's reasons is at best unnecessary and at worst irrational. It has been consistently rejected by Isaac Levi (e.g., Levi 1980) and does not figure in the well-known AGM theory (Alchourrón, Gärdenfors, and Makinson 1985). For a recent expression of Levi's aversion toward what he calls "pedigree epistemology", see his 2004 book, e.g., pp. 231-32.

[20] The author wishes to thank Wlodek Rabinowicz for his useful comments on an earlier version. The main ideas were discussed at the Higher Seminar in Theoretical Philosophy in Lund. The author is indebted to the participants, in particular to Staffan Angere, Anna-Sofia Maurin and Niklas Vareman, for their input. Finally, the author benefited greatly from the many suggestions for improvements that he received from the Editor, Neil Tennant.