Knowledge first, stability and value

Barnaby Walker[1]

Abstract

What should knowledge first theorists say about the value of knowledge? In this paper I approach this issue by arguing for a single 'modest knowledge first claim' (MKF) about the value of knowledge. This is that the special value of knowledge isn't merely instrumental value relative to true belief. I show that MKF is inconsistent with the version of the Platonic stability theory that Williamson defends in *Knowledge and its Limits*. I then argue in favour of MKF by arguing that Williamson's stability theory fails for reasons that plausibly generalise to any theory of the value of knowledge that is inconsistent with MKF. Crucial to this argument is a putative adequacy condition on philosophical theories of the value of knowledge: that, in order to fully explain the special value of knowledge, a theory must identify respects in which knowledge is more valuable than true belief that give enquirers adequate reason for preferring knowledge to true belief. I conclude by suggesting that this adequacy condition is the source of a more general dilemma for proponents of philosophical theories of the value of knowledge.

Keywords Value of knowledge · Knowledge first approach · Stability theory · Meno problem · Enquiry

In the preface to *Knowledge and its Limits* Williamson tells us that the knowledge first theorist 'takes the simple distinction between knowledge and ignorance as a starting point from which to explain other things, not as something itself to be explained' (2000: v). This methodological principle applies straightforwardly to issues about the nature of knowledge, belief, justification, evidence, and so on: it tells us that instead of trying to explain what knowledge is using notions like belief and justification, we should try using the notion of knowledge to explain things like justification and belief. However, whether this principle can be developed in ways that are germane to issues about the value of knowledge is less obvious, and knowledge first theorists have not

✉ Barnaby Walker
  B.Walker.3@warwick.ac.uk

1 Department of Philosophy, Social Sciences Building, University of Warwick, Coventry CV4 7AL, UK

Springer

been forthcoming with the details. What should knowledge first theorists say about the value of knowledge?

This paper approaches this large question by focusing on a single modest thesis about the value of knowledge. This thesis, which I refer to as 'the minimal knowledge first value claim' (or 'MKF' for short), says that the special value of knowledge is not merely instrumental value relative to true belief. Consequently, we cannot fully explain the special value of knowledge merely by identifying distinctive ways in which knowledge is instrumental to true belief. Though modest, the thesis is carefully chosen: close inspection reveals it to be inconsistent with the account of the value of knowledge that Williamson presents in *Knowledge and its Limits*, provided we construe that account as an attempt to fully explain the special value of knowledge. My argument for MKF is somewhat indirect. I begin by analysing a well-known objection to the stability theory, that it cannot explain why knowledge is more valuable than true belief in cases in which true belief lacks a 'shelf life'. Reflection on this objection reveals a more general adequacy condition on philosophical theories of the value of knowledge. This is that, in order to fully explain the special value of knowledge, a theory must identify respects in which knowledge is more valuable than true belief that give enquirers adequate reason for preferring knowledge to true belief across the full range of circumstances in which they want to know. I then argue that Williamson's stability theory fails to identify such a respect, and that it does so for reasons that plausibly generalise to any theory of the value of knowledge upon which the special value of knowledge is merely instrumental value relative to true belief. Thus, I suggest, powerful (though not decisive) grounds for accepting MKF emerge from reflection on the shortcomings of Williamson's stability theory.

I begin, in Sect. 1, by identifying several knowledge first claims about the value of knowledge and clarifying the claim, MKF, on which I shall focus. Section 2 outlines the theory of the value of knowledge that Williamson offers in *Knowledge and its Limits*, which is a version of the Platonic idea that knowledge, unlike true belief, is stable. Section 3 highlights the inconsistency between Williamson's account and MKF by showing that it is precisely because Williamson's account explains the special value of knowledge by appealing to a distinctive way in which knowledge is instrumental to true belief that it is able to avoid an objection offered by Jonathan Kvanvig (2003). Together, Sects. 2–3 clarify both the motivation for, and the details of, Williamson's account, including correcting a minor defect in Williamson's case for his probabilistic stability claim. In Sect. 4 I present my argument for MKF, based on considerations that emerge from the objection that the stability theory cannot explain the value of knowledge in circumstances in which true belief lacks a shelf life. I conclude, in Sect. 5, by briefly considering the larger implications of my argument for philosophical debates about the value of knowledge.

1 A knowledge first claim about value

In considering the implications of knowledge first epistemology for philosophical issues about the value of knowledge, a helpful starting point is provided by the methodological principle Williamson proclaims at the outset of *Knowledge and its Limits*. To

repeat, this is the principle that the knowledge first theorist 'takes the simple distinction between knowledge and ignorance as a starting point from which to explain other things, not as something itself to be explained' (2000: v). Williamson's methodological principle has positive and negative elements. Negatively, the principle says that we are not to explain what knowledge is in terms of other things; positively, it says that we should instead explain other things in terms of knowledge. Both elements of Williamson's methodological principle are open-ended, and both are developed further in his book. In denying that we can explain knowledge in terms of other things, Williamson is primarily concerned to deny that we can conceptually analyse knowledge as a conjunction of belief and truth with further factors. He allows that an explanation of the concept of knowledge can be given in terms of the notion of a factive mental state operator (FMSO).[1] Regarding the positive component of his principle, Williamson offers accounts of belief, evidence and assertion in terms of knowledge.[2]

For present purposes, the significance of Williamson's methodological principle is that we can construct parallel positive and negative claims about the value of knowledge. The parallel negative claim would be that we cannot exhaustively explain the value of knowledge in terms of the value of other things; the parallel positive claim would be that we should instead explain the value of other things in terms of the value of knowledge. Like the positive and negative elements of Williamson's original principle, each of these claims is open-ended and can be developed in a number of ways. Depending on how we further articulate them, we arrive at stronger or weaker knowledge first claims about the value of knowledge.[3]

My aim here is to defend a negative knowledge first claim about the value of knowledge, that is, a version of the claim that we cannot exhaustively explain the value of knowledge in terms of the value of other things. The strongest way of developing such a claim would be to insist that knowledge is intrinsically valuable, or valuable for its own sake. If knowledge is intrinsically valuable, then its value cannot be fully explained in terms of the value of other things, although it is of course consistent with this idea that the value of knowledge is partly instrumental. One needn't commit oneself to the idea that knowledge is intrinsically valuable to defend a negative knowledge first claim about value, however. Pritchard (2010) speculates that Williamson would be attracted to a thesis Pritchard calls 'Epistemic Value K-Monism':

> Knowledge is the sole fundamental epistemic good. (12)

In Pritchard's terms, a fundamental epistemic good is an epistemic good whose epistemic value is at least sometimes not merely instrumental value relative to a further epistemic good. In other words, it is an epistemic good that is at least sometimes epistemically valuable entirely for its own sake. An epistemic good is non-fundamental

---

[1] See Williamson (2000: pp. 27–41). For further discussion of Williamson's explanation of the concept of knowledge in terms of the notion of an FMSO, and its relation to the traditional programme of analysing knowledge, see Cassam (2009) and Williamson's response in the same volume.

[2] More recently, Sutton (2007) has extended the positive component of the knowledge first programme to justification.

[3] Note that there is no suggestion that the knowledge first value claims I discuss are entailed by Williamson's methodological principle. The reason for counting them as knowledge first claims lies merely in the structural parallel between them and knowledge first claims about the nature of knowledge.

just in case it is not a fundamental epistemic good (see 2010: pp. 11–12). Importantly, according to Pritchard, it doesn't follow from the fact that something is a fundamental epistemic good that it is intrinsically valuable. This is because it is consistent with the fact that something is a fundamental *epistemic* good that its value is instrumental value relative to further *non-epistemic* goods. Thus, although Epistemic Value K-Monism doesn't commit one to the claim that knowledge is intrinsically valuable, it does commit one to the claim that we cannot explain the epistemic value of knowledge by identifying ways in which knowledge is instrumental to further epistemic goods. Epistemic Value K-Monism combines this negative knowledge first claim with a positive one: the epistemic value of all epistemic goods besides knowledge is to be explained by showing how those goods are instrumental to knowledge. This is the point of insisting that knowledge is the *sole* fundamental epistemic good.

The negative knowledge first claim I will defend here is distinct from both Epistemic Value K-Monism and the claim that knowledge is intrinsically valuable. It says that the special value of knowledge isn't purely instrumental value relative to true belief. Consequently, we cannot fully explain the special value of knowledge by identifying distinctive ways in which knowledge promotes true belief. I refer to this claim as the 'minimal knowledge first value claim', or 'MKF' for short.

MKF specifically concerns the *special* value of knowledge. By this I mean the value knowledge has over and above true belief, and also over and above states intermediate between true belief and knowledge. State S is intermediate between true belief and knowledge, in the present sense, just in case (i) that one knows p entails that one is in S, (ii) that one is in S doesn't entail that one knows p, (iii) that one is in S entails that one truly believes that p, and (iv) that one truly believes that p doesn't entail that one is in S.[4] The obvious putative example of such an intermediate state is that of having the justified true belief that p. The problem of explaining why knowledge is more valuable than states intermediate between true belief and knowledge is known as the 'secondary value problem', in contrast to the 'primary value problem', posed by Meno, of explaining why knowledge is more valuable than true belief.[5] Explaining the special value of knowledge in the present sense is therefore a matter of resolving both the Meno (or primary value) and secondary value problems. Note that in referring to the value knowledge has over and above both true belief and states intermediate between true belief and knowledge as the 'special' value of knowledge, I do not intend to suggest anything about the way in which knowledge is more valuable than these other states. For example, I am not presupposing that the greater value of knowledge is due to a difference in the kind, and not merely degree, of value that knowledge possesses.[6]

---

[4] Note that, on this understanding of what it is for a state to be intermediate between true belief and knowledge, being in an intermediate state doesn't exclude one from being in the state of knowing. Indeed, given this definition, if one knows that p then one must be in any state intermediate between knowing p and truly believing p. All that follows from the fact that one knows p is that one isn't *merely* in a state intermediate between knowledge and true belief. (Thanks to an anonymous referee for seeking clarification of this point.).

[5] These labels for the problems are due to Pritchard (2007), but the formulation of the secondary value problem in terms of the notion of an intermediate state is my own.

[6] Some philosophers do take the greater value of knowledge to be a difference of kind and not merely of degree. For example, in addition to the primary and secondary value problems, Pritchard (2007) distinguishes

The point of calling MKF a 'minimal' knowledge first claim is to emphasise that the claim it makes is heavily restricted. It restricts the open-ended claim that we cannot exhaustively explain the value of knowledge in terms of the value of other things in three ways: (i) it interprets this claim as denying only that the value of knowledge can be fully explained by showing how knowledge is *instrumentally valuable* to other things; (ii) it restricts this denial to *true belief* alone; and (iii) it denies only that we can *fully* explain the *special value* of knowledge by showing how knowledge is instrumental to true belief, where the 'special value' of knowledge is understood, as above, to mean the value that knowledge has over and above true belief, and also over and above states intermediate between true belief and knowledge. Unlike Pritchard's Epistemic Value K-Monism, MKF concerns value in general rather than epistemic value specifically, but, like Epistemic Value K-Monism, it doesn't entail that knowledge is intrinsically valuable.[7] If MKF is false, then the prospects for defending a knowledge first approach to the value of knowledge seem dim.

Although there has been relatively little discussion of knowledge first claims about the value of knowledge in recent epistemology, there has been ample discussion of knowledge first claims about epistemic norms. For example, Williamson (2000) argues that the act of assertion is partly constituted by the norm that one must: assert p only if one knows p. Knowledge norms have also been proposed for action and belief.[8] Discussions of epistemic norms typically take place in isolation from debates about the value of knowledge, but it might be thought that the idea that knowledge is a normative condition of assertion, belief or action can be used to explain the value of knowledge. For example, if knowledge and not merely true belief is a normative condition of action, then it might be thought that this helps to explain why knowledge is more valuable than true belief. This is an intriguing suggestion, but I will not pursue it further here. As Dutant (2014) points out, that knowing p is a normative condition of belief, assertion, or action doesn't entail that knowledge is valuable, let alone that it is more valuable than true belief. It might be possible to construct an argument for the claim that knowledge is more valuable than true belief by supplementing the claim that knowledge is a normative condition of belief, assertion, or action with further premises linking claims about epistemic norms with claims about value. However, the investigation of this type of argument is beyond the scope of the present paper.[9]

Footnote 6 continued

a further tertiary value problem. This is the problem of explaining why knowledge is more valuable than true belief and other intermediate states, not merely as a matter of degree, but in kind (104, n. 4). I will not discuss the tertiary value problem further here.

[7] The logical relationship between MKF and Epistemic Value K-Monism is more complicated. Certainly, MKF doesn't entail Epistemic Value K-Monism; for example, it is consistent with MKF that there are fundamental epistemic goods besides knowledge. Less obviously, the converse entailment also fails: even if knowledge is the sole fundamental epistemic good, it might still be the case that the special value of knowledge is purely instrumental value relative to true belief. This is because, consistently with the specifically epistemic value of true belief being merely instrumental value relative to the epistemic value of knowledge, it might still be the case that the value of knowledge, including its epistemic value, is to be explained instrumentally relative to the *practical* value of true belief.

[8] For discussion of knowledge norms for action and/or belief, see (e.g.) Williamson (2018), Hawthorne and Stanley (2008), and Gibbons (2013).

[9] For further discussion of the prospects of such an argument, see Dutant (2014).

As far as I know, MKF is not a claim that anyone has explicitly advocated or denied.[10] However, it doesn't follow from this that it is devoid of philosophical interest. Surprisingly, MKF is inconsistent with the version of the stability theory of the value of knowledge that Williamson defends in *Knowledge and its Limits*, provided we construe that theory as an attempt to fully explain the special value of knowledge. The next two sections explain why.

2 Williamson's stability theory

The stability theory is normally expressed as the claim that knowledge is more valuable than true belief because knowledge is more stable, i.e. less easily lost, than true belief. The theory was originally proposed by Plato in the *Meno* (Plato 2005). In a famous analogy, Plato likens the difference between knowledge and true belief to the difference between tethered and untethered statues of Daedalus, a mythical sculptor whose productions were so lifelike that they ran away unless they were tied down. It is on account of its being 'tied down', and thus stable—a metaphor that Plato cashes out in terms one's working out the reason why one's belief is true (98a), and ultimately in terms of his doctrine of recollection—that knowledge is more valuable than true belief.[11] An initially attractive feature of the stability theory is that it allows that, in a certain way, knowledge really is no more valuable than true belief: other things being equal, someone who knows that *p* over some interval of time is no better off *during that interval* than they would have been if they had merely truly believed that *p* over that interval. If one is taken by this thought, as I think many are when first confronted with the Meno problem, then it seems that the superiority of knowledge can only lie in its being somehow more durable than true belief. The stability theory is apt to appear compulsory.[12]

---

[10] Pritchard and Turri (2017) claim that the swamping problem will affect any theory of the value of knowledge 'which regards the greater value of knowledge over true belief as instrumental value, where the instrumental value in question is relative to the valuable good of true belief', but do not endorse MKF for this reason. However, it is false that the swamping problem, as traditionally conceived, affects any theory of the value of knowledge implying that the special value of knowledge is merely instrumental value relative to true belief. The intuition behind the traditional conception of the swamping problem is that, where some good (e.g. true belief) is already present, the presence of something else that is merely instrumentally valuable to the acquisition of that good (e.g. a true belief's having been produced by a reliable method) adds no further value. The value of the good 'swamps' the value of the thing that is instrumental to its acquisition. However, this problem won't arise for a theory of the value of knowledge according to which knowledge is more valuable than true belief because knowledge has greater instrumental value relative to *future* true belief. On such a theory, knowledge isn't (merely) instrumental to the *acquisition* of true belief, but to its *retention*; consequently, the value of knowledge isn't 'swamped' by the value of *present* true belief. As we will see in Sects. 2–3, Williamson's version of the stability theory of the value of knowledge is a theory of precisely this kind. (Thanks to an anonymous referee for pressing me on the relationship between MKF and the swamping problem.).

[11] In attributing this theory to Plato I am of course assuming that the views expressed through the character of Socrates in the relevant part of the *Meno* are Plato's own. Socrates later admits that he doesn't take himself to know that the theory of the value of knowledge he offers is correct (98b).

[12] This thought is arguably part of what motivates Plato's (provisional) acceptance of the theory. Socrates stresses that 'As long as they stay put, true beliefs too constitute a thing of beauty and do nothing but good' (*Meno* 97e–98a).

In approaching Williamson's stability proposal it is helpful to begin by recalling some remarks Bernard Williams once made about Plato's version of the theory. The first remark is a negative one. A defence of the claim that knowledge is more stable than true belief should not be made on the basis of the 'very dubious' and 'blankly psychological' assertion that 'one is more disposed to forget what one merely believes than what one knows' (2005: p. 24). Although we might speculate that the processes by which subjects acquire certain kinds of knowledge render them unlikely to forget knowledge so acquired—e.g. that learning a theorem of geometry by constructing a proof of it, as Meno's slave boy does in the dialogue, has the effect of cementing that theorem in memory—it is a familiar fact that there also ways of lastingly imprinting beliefs that are not ways of acquiring knowledge (e.g. indoctrination). Williams's second remark is positive, and suggests an alternative. Rather than taking Plato to be advancing his proposal on the basis of a premise about forgetting, we should instead take him to be making the point, 'more interesting for the theory of knowledge, that knowledge cannot rationally be rendered doubtful' (2005: p. 24). Even if knowledge is no less likely to be forgotten than true belief, knowledge is more stable than true belief because there is a way in which true beliefs can be lost that knowledge cannot: namely, by being rationally undermined by future evidence that comes into the mere true believer's possession.

Although I think we should agree with Williams that considerations about memory are not a promising basis for the stability theory, the alternative he suggests is also dubious. The claim that knowledge cannot rationally be rendered doubtful seems too strong: *prima facie*, it isn't difficult to imagine cases in which someone who knows that *p* is rationally led to doubt that *p*, and thus to lose her knowledge that *p*.[13] However, it would be an error to conclude from this that the stability theory cannot be defended on the basis of considerations about the evidential robustness of knowledge. To establish the stability theory on the basis of such considerations, it isn't necessary to appeal to the strong claim that knowledge *cannot* rationally be rendered doubtful: we need only invoke the weaker principle that knowledge is *less likely* than true belief to be rationally undermined by future evidence.

This is precisely the move that Williamson makes.[14] He formulates his key claim as follows:

> Present knowledge is less vulnerable than mere present true belief to rational undermining by future evidence, *which is not to say that it is completely invulnerable to such undermining* [my emphasis]. (2000: p. 79)

Call this claim 'the evidential robustness of knowledge thesis', or 'ERK' for short. Although ERK is unscathed by the possibility of a subject's knowledge being rationally undermined by future evidence that comes into her possession, it makes the still

---

[13] For a plausible example of this kind see Kripke (2011: pp. 35–36). Note that I do not mean to suggest that Williams himself would have accepted the principle that knowledge cannot rationally be rendered doubtful. Some of his subsequent remarks in *Descartes* suggest that he would not—in particular, his disavowal of the idea that our everyday concept of knowledge requires one to be specially certain of what one knows, or even to have conscious reasons for believing what one knows (2005: p. 27).

[14] Williamson (2000: p. 78, n. 4) notes Williams's earlier discussion. It seems likely that Williamson was led to his own version of the stability theory by reflecting on the points Williams makes.

significant assertion that mere true belief is more vulnerable to being lost in this way than knowledge. In support of this assertion, Williamson identifies two ways in which a subject who truly believes that *p* can fail to know that *p* which render her belief that *p* peculiarly vulnerable to being rationally undermined by future evidence. The first is that mere true belief, unlike knowledge, can be 'essentially based' (2000: p. 78) on false beliefs. Thus, my true belief that Exeter is in the south west of England might be based on my false belief that Exeter is in Cornwall and my (true) belief that Cornwall is in the south west, but knowledge that Exeter is in the south west of England cannot be so based. Consequently, the potential for discovering the falsity of other beliefs on which one's mere true belief is essentially based is one way in which mere true belief, but not knowledge, can be vulnerable to rational undermining by future evidence. The second is that mere true belief, unlike knowledge, can exist in circumstances in which misleading evidence against the truth of the belief is rife (Williamson 2000: pp. 78–79). Since misleading evidence against the truth of the belief is rife in the believer's environment, her belief is especially vulnerable to being rationally undermined.[15]

Having supported ERK, Williamson takes it to establish a version of the stability proposal:

> If your cognitive faculties are in good order, the probability of your believing *p* tomorrow is greater conditional on your knowing *p* today than on your merely believing *p* truly today (that is, believing *p* truly without knowing *p*). Consequently, the probability of your believing *p* tomorrow is greater conditional on your knowing *p* today than on your believing *p* truly today. (2000: p. 79)[16]

Note that Williamson's version of the stability claim is conditional on the subject's cognitive faculties being in 'good order'. This restriction is necessary, according to Williamson, because

> …profoundly dogmatic beliefs which are impervious to future evidence and do not constitute knowledge may be even more likely to persist than beliefs that are rationally sensitive to future evidence and do constitute knowledge. (2000: p. 79)

By limiting the stability claim to subjects whose cognitive faculties are in good order, Williamson excludes such profoundly dogmatic beliefs from the calculation of the relevant probabilities (viz. the probability of a subject believing *p* tomorrow conditional

---

[15] Obviously, the specific claims on which Williamson bases his defence of ERK—that knowledge cannot be 'essentially based' on false beliefs, or exist in environments rife with misleading counterevidence—are controversial, and at least require further clarification. (Exactly what is it for a belief to be 'essentially based' on false beliefs? What constitutes a believer's environment? And so on.) However, the general point that a subject who believes that *p* can fail to know that *p* for reasons that render her belief peculiarly vulnerable to being rationally undermined by future evidence is immensely plausible. Furthermore, the critical points I wish to make about Williamson's stability theory in no way depend on the details of Williamson's defence of ERK. For these reasons, I will not discuss these issues further here.

[16] Williamson notes that the inference in this passage depends on the principle governing conditional probabilities that 'if E entails D and $P[C|D \wedge \sim E] < P[C|E]$ and $P[E|D] < 1$, then $P[C|D] < P[C|E]$' (2000: p. 79, n. 7).

on her knowing p today versus the probability of a subject believing p tomorrow conditional on her merely believing p truly today), because the cognitive faculties of a subject who harbours profoundly dogmatic beliefs aren't in good order. Thus, the point that stubbornly dogmatic mere true beliefs can be more stable than knowledge isn't an objection to Williamson's stability theory.[17]

We appear to have a solution to the Meno problem. Provided her cognitive faculties are in good order, a subject who knows that p is more likely to believe that p in the future, other things being equal, than a subject who truly believes that p. Thus, if having true beliefs in the future is good, knowledge is more valuable than true belief. The next question to consider is whether Williamson's theory is consistent with the minimal knowledge first value claim described in Sect. 1.

3 Williamson's stability theory and the minimal knowledge first value claim

It might seem that nothing more needs to be said on this score. A moment ago I quoted Williamson's probabilistic version of the stability claim, and that formulation unambiguously states that what is more probable in the future conditional on your now knowing p is your *believing* that p. Clearly, then, Williamson's version of the stability theory explains why knowledge is more valuable than true belief by appealing to the idea that knowledge has greater instrumental value relative to future true belief than true belief simpliciter. However, MKF says that the special value of knowledge isn't merely instrumental value relative to true belief. Taken as an attempt to provide a full explanation of why knowledge is more valuable than true belief,[18] Williamson's theory is therefore straightforwardly inconsistent with MKF. Although it shouldn't be controversial that Williamson's theory purports to explain the special value of knowledge by identifying a distinctive way in which knowledge is instrumental to future true belief, this point has been missed in discussions of Williamson's theory. To illustrate this misunderstanding and further clarify Williamson's view, I will show how it undermines a superficially plausible objection to Williamson's theory due to Jonathan Kvanvig (2003).

Kvanvig's objection targets Williamson's case for ERK.[19] As we saw earlier, Williamson supports ERK—i.e. the claim that present knowledge is less vulnerable

---

[17] However, in restricting his stability claim in this way, Williamson opens himself up to another line of objection—namely, one arising from concerns about the extent to which our cognitive faculties are in good order. On the face of it, the fact that the true beliefs of a subject who knows are more stable if her cognitive faculties are in good order cannot explain why knowledge is more valuable than true belief *for us* unless *we* are subjects whose cognitive faculties are in good order. For a discussion of the extent to which this is true of our cognitive faculties, see Cassam (2014: Chapter 2); especially relevant here is the phenomenon of belief perseverance.

[18] It should be said that it isn't clear that Williamson intends his theory to be taken in this way. His elaboration of the stability theory is a minor part of *Knowledge and its Limits*, and occurs in a chapter in which his primary concern is to defend the claim that knowledge is a mental state by defending the claim that knowledge plays an ineliminable role in the causal explanation of action.

[19] An underlying issue with Kvanvig's discussion is that he misconstrues the relationship between Williamson's probabilistic stability claim and ERK: he presents the former as an attempt to precisify the latter (see, e.g., p. 14), whereas on my reading they are distinct claims, and ERK is one of the premises

than present mere true belief to rational undermining by future evidence—by identifying ways in which a subject can truly believe that *p* without knowing that *p*, where those ways render the subject's belief peculiarly liable to being rationally undermined by future evidence. However, even someone who accepts the points that Williamson makes about this might still question whether he has done enough to establish ERK. That there are ways in which mere true belief but not knowledge can be rationally undermined by future evidence shows that knowledge is more robust in the face of future evidence only if there are *not* also ways in which knowledge but not mere true belief can be rationally undermined by future evidence. Williamson provides no argument for this further claim, and Kvanvig identifies a plausible reason for denying it.

Kvanvig's reason exploits a familiar idea about knowledge. This is that the presence of misleading counterevidence that *not-p* in the environment of a subject who truly believes that *p* can prevent the subject from knowing that *p*, even if the subject herself is unaware of that counterevidence.[20] If one is persuaded by this idea, one is likely to think, furthermore, that a subject can *lose* her knowledge that *p* as a consequence of such evidence *coming to exist* in her environment, subsequently to her original acquisition of knowledge. Kvanvig illustrates this point using a mathematical example:

> …my mathematical knowledge might be undermined tomorrow by the sincere testimony of a renowned mathematician to the effect that what I believe is false. Until such testimony is rendered, I have such knowledge, but I lose it when the defeating testimony is given, even though I am unaware that such testimony has occurred. (2003: p. 15)

As Kvanvig goes on to observe, in this case the testimony of the renowned mathematician doesn't undermine my belief in the mathematical proposition. The fact that a subject can lose her knowledge that *p* as a consequence of misleading counterevidence that *not-p* coming to exist in her environment, even if she is unaware of that counterevidence, calls ERK into question, for it shows that there is a way in which knowledge is vulnerable to rational undermining by future evidence that mere true belief is not. One might think, therefore, that whether ERK is true or not depends on the nature of the subject's environment. This is what Kvanvig suggests:

> …there will be worlds in which known beliefs are more susceptible to cessation than are mere true beliefs. Such susceptibility would occur when (i) the unpossessed information that undermines knowledge for some true beliefs is resistant to discovery, while (ii) the universe conspires to generate future defeaters for a large percentage of what is known. So whether Williamson's inequality [i.e. probabilistic version of the stability claim] is true or not depends on what kind of world we live in and is thus a contingent truth at best. (2003: p. 17)

Footnote 19 continued

from which Williamson derives the probabilistic stability claim as a conclusion. Given this misconstrual, it is understandable why Kvanvig takes himself to be attacking Williamson's stability theory in attacking ERK. I am convinced that a close reading of the relevant passage confirms the interpretation of Williamson's theory presented in Sect. 2, however.

[20] The famous example in this regard is Harman's dictator case (see his 1973: pp. 143–144). As I pointed out earlier, Williamson himself appeals to this idea in arguing for ERK.

According to Kvanvig, the contingency of Williamson's inequality means that it cannot explain why knowledge is more valuable than true belief, because

> …the value of knowledge does not covary with the truth value of Williamson's inequality. It is simply false that knowledge loses its value in worlds where the environment is less cooperative… (2003: p. 17)

How strongly does this objection speak against the idea that we can explain the value of knowledge in terms of stability?

One response to the objection would be to question Kvanvig's assumption that a theory of the value of knowledge must explain why knowledge is *necessarily* more valuable than true belief.[21] It isn't obvious that the intuition that knowledge is more valuable than true belief includes the idea that knowledge is necessarily more valuable than true belief. However, there is in fact a much more clear cut problem with Kvanvig's argument.

As I have emphasised, Williamson's stability claim is that if a subject's cognitive faculties are in good order the probability of her believing *p* tomorrow is greater conditional on her knowing *p* today than it is on her merely believing *p* truly today. *That knowledge but not belief can be lost as a consequence of rational undermining by future evidence of which the subject is unaware has no tendency to undermine this claim*. In cases in which a subject loses her knowledge that *p* because of evidence that *not-p* of which she is unaware, the subject doesn't thereby lose her belief that *p*. Thus, such cases do not speak against the probabilistic inequality claim Williamson defends, and cannot be used to argue that it is contingent on the subject's being in a 'cooperative' environment. They do, of course, speak against the alternative probabilistic inequality claim that the probability of a subject *knowing p* tomorrow conditional on her knowing *p* today is greater than the probability of a subject believing *p* tomorrow conditional on her truly believing *p* today, but Williamson nowhere advocates this claim.[22]

The point that Williamson's version of the stability theory is concerned only with the stability of one's beliefs when one knows tends to be occluded by two misleading features of the debate. The first is the way in which Williamson argues for his stability claim. As I explained earlier, Williamson formulates ERK as the thesis that 'Present knowledge is less vulnerable than mere present true belief to rational undermining by future evidence' (2000: p. 79). This thesis arguably is contingent on the extent to which a subject's environment conspires to generate future hidden defeaters for her knowledge. However, it should be clear that this possibility is irrelevant to the probabilistic inequality claim that Williamson defends on the basis of ERK, and consequently that, for the purposes of establishing that claim, Williamson need only appeal

---

[21] As suggested by Fricker (2009: p. 131) and Olsson (2007: pp. 350–351). Millar (2011: p. 71, n. 12) downplays the significance of cases in which a subject loses her knowledge without losing her true belief, claiming that, even if such cases can occur, 'instances are surely rare'.

[22] Note that the present point is independent of the specific way that Kvanvig alleges a subject can lose her knowledge that *p* without thereby losing her belief that *p*. Quite generally, the claim that the probability of subject believing *p* in the future is greater conditional on her knowing that *p* than on her truly believing that *p* isn't undermined by identifying ways in which a subject can lose her knowledge that *p* without losing her belief that *p*, because even when such ways are operative the subject retains her belief that *p*. Thus, there is no possibility of rescuing Kvanvig's objection by specifying further ways in which a subject can lose her knowledge that *p* without losing her belief that *p*.

to the premise that when a subject knows that *p her belief* that *p* is less vulnerable to rational undermining by future evidence than the mere true belief that *p*. This premise isn't contingent on the extent to which a subject's environment generates future hidden defeaters for her knowledge. Some of the blame for the misunderstanding of his stability theory must therefore be apportioned to Williamson himself.

The other misleading feature of the debate is that the stability theory is normally formulated as the claim that *knowledge* is more stable than true belief, regardless of the specific version of the theory under discussion. (This, indeed, was how I introduced the theory in the previous section.) Strictly speaking, however, Williamson only advocates a claim about the greater stability of one's *beliefs* when one knows. When one is theorising about ways in which knowledge might be more valuable than true belief, or than other states intermediate between true belief and knowledge, it is important to attend to whether the beneficial features associated with knowledge are features of knowledge itself or only features of some state implicated by knowledge. As Kvanvig's objection shows, failure to attend to this point can lead to misplaced concerns.

Let me summarise the argument so far. I began by introducing the minimal knowledge first value claim (Sect. 1). To repeat, this was that the special value of knowledge isn't merely instrumental value relative to true belief, and consequently that we cannot fully explain the special value of knowledge merely by identifying distinctive ways in which knowledge promotes true belief. I then outlined Williamson's stability theory (Sect. 2). I pointed out that Williamson's stability theory is inconsistent with MKF, because it purports to explain why knowledge is more valuable than true belief by identifying a way in which knowledge better promotes future true belief than true belief *simpliciter*. I then showed how overlooking this aspect of Williamson's position generates a misguided objection to his theory (Sect. 3).

The next section introduces a further objection to the stability theory. This is that the stability theory is unable to explain the special value of knowledge in cases in which true belief lacks a 'shelf life'. As we will see, reflecting on this objection unearths a more general consideration that tells against any theory of the value of knowledge upon which the special value of knowledge is merely instrumental value relative to true belief.

4 True beliefs without a shelf life

Craig (1990) objects to the stability theory on the grounds that it cannot explain the special value of knowledge in cases where true belief lacks a 'shelf life'.[23] Here is how Craig puts the argument:

> Whether the stabilisation of true beliefs is important or not depends on which beliefs we are considering, and the circumstances of the agent—many beliefs are required for the guidance of single, 'one-off' actions under circumstances which will not recur, and once the particular occasion is past there is no obvious value at all in their persistence. (I might now need a true belief about the time;

---

[23] The shelf life metaphor is due to Hyman (2010), but the earliest source for the objection I know of is Craig (1990).

but that this belief should persist, so that tomorrow I will still know what the time was today, at the moment when I wanted to know it, may be of no interest to me whatever.) (1990: p. 7)

The crux of Craig's objection is straightforward: there are circumstances in which the stability of a true belief makes it no more valuable, but in which knowledge remains more valuable than true belief; hence, it cannot be purely in virtue of the greater stability of one's beliefs when one knows that knowledge is more valuable than true belief. Craig's reason for thinking that there are circumstances in which the stability of a true belief makes it no more valuable is that the stability of a true belief makes it more valuable only if there is value in the persistence of that true belief (i.e. only if true belief 'has a shelf life'), but the persistence of a true belief often has no practical value for the agent, because it won't guide any of her future actions. Consequently, Craig thinks, there is 'no obvious value at all' in the persistence of many true beliefs. This argument can clearly be questioned, but I want to begin by considering a different issue about the objection as Craig presents it. This is why we should accept the other implicit premise of his argument—that knowledge is more valuable than true belief even when true belief lacks a shelf life.

Craig offers no explicit argument for this premise. It isn't clear that appealing to the existence of a general intuition that knowledge is more valuable than true belief is sufficient to justify the claim that knowledge is more valuable than true belief in cases where true belief lacks a shelf life. As Pritchard and Turri (2017) point out, the assertion that knowledge is more valuable than true belief is implausible when construed as a strict generalisation—i.e. as the claim that *any* instance of propositional knowledge is more valuable than the corresponding instance of mere true belief.[24] Consequently, the reason for thinking that knowledge is more valuable than true belief in cases where true belief lacks a shelf life cannot be that *any* instance of knowledge is more valuable than the corresponding instance of mere true belief. However, if we construe the content of the intuition that knowledge is more valuable than true belief as merely generic, akin to the content of 'ducks lay eggs', then it's not clear that it provides us with grounds for thinking that knowledge is more valuable than true belief when true belief lacks a shelf life. Generic statements permit exceptions, so it might be argued that cases in which true belief lacks a shelf life are simply amongst the exceptions to the generalisation that knowledge is more valuable than true belief. If the shelf life objection is to seriously threaten the stability theory, further support is required for the idea that knowledge is more valuable than true belief when true belief lacks a shelf life.

At this juncture it is instructive to reflect on the example Craig uses to illustrate his objection. This example is one in which I need a true belief about the present time, but have no interest in the persistence of that true belief. Reflecting on this example is instructive because it reveals a line of thought that supports the idea that knowledge is more valuable than true belief even when true belief lacks a shelf life.

In a situation in which I need at least a true belief about the present time, what is the natural way for me to express my desire to have the truth about the time? The

---

[24] Pritchard and Turri's counterexample to the strict generalisation is a situation in which knowing that *p* would kill you but merely truly believing that *p* would win you the lottery.

natural way, as Craig acknowledges, is for me to say that I want to *know* what the time is, and this is so irrespective of whether there is any practical value for me in the persistence of a true belief about the time. (It would be odd, in ordinary circumstances, to say that I wanted a true belief about the time, or to ask the indirect question, 'Do you have a true belief about the time?'.) Now, if I want to know what the time is, it seems reasonable to ask why, in particular, I want to *know* this. Why do I want to know it, rather than (e.g.) merely wanting to have a true belief about it? An intuitive response to this question is that I would *prefer* knowing the time to merely having a true belief about it, but this can hardly be the end of the matter. Although there are such things as personal or idiosyncratic preferences (e.g. preferring brown sauce to ketchup on bacon sandwiches), the preference for knowledge over true belief surely isn't one of them, for this preference appears to be universal.[25] This suggests that there are reasons for preferring knowledge to true belief that aren't idiosyncratic, or, at least, that we take there to be such reasons. In other words, it suggests that we *value* knowledge more highly than true belief. That we do so is evidence that knowledge really is more valuable than true belief. Admittedly, this evidence isn't decisive—our valuations aren't infallible. However, in the absence of an account of how we might have fallen into error, the failure of a theory to explain why knowledge is more valuable than true belief in circumstances in which we want to know is a reason to think that the theory is inadequate. The fact (if it is a fact) that we want to know the answers to questions in circumstances in which true beliefs about the answers to those questions would lack a shelf life is therefore a reason to think that the stability theory is inadequate as it stands.[26] The present point isn't that this reasoning—from desiring, through preferring, to valuing and value—is incontestable, or shouldn't be contested. (I will briefly discuss the implications of rejecting the reasoning in the next section.) It is merely to identify potential grounds for thinking that knowledge is more valuable than true belief even when true belief lacks a shelf life.

One effect of making these grounds explicit is to highlight the fact that the shelf life objection is merely one version of a much more general way of challenging philosophical theories of the value of knowledge. To challenge a theory of the value of knowledge in this way, one first identifies a type of situation on which, on the theory

---

[25] Pertinent here is Craig's observation that 'There seems to be no known language in which sentences using 'know' do not find a comfortable and colloquial equivalent' (1990: p. 2).

[26] An anonymous reviewer has suggested to me that, irrespective of the merits of this way of developing the shelf life objection, it isn't a way of developing the objection that Craig would accept. The starting point for Craig's account of the concept of knowledge is an enquirer who merely wants true beliefs about the answers to questions (see 1990: p. 11). Thus, the reviewer suggests, Craig wouldn't be happy with the idea that, in many ordinary contexts (including contexts in which true belief lacks a shelf life), we want knowledge and not just true beliefs. Since my aim here is just to develop the shelf life objection in the most powerful way I can, it doesn't matter for my purposes if the way I develop the objection is in tension with other views Craig holds. However, I should also say that it isn't at all clear to me that there is any tension between Craig's views and my way of developing the objection. In taking an enquirer who merely wants true beliefs as the starting point for his genealogical investigation, Craig doesn't commit himself to the claim that ordinary human beings, as things actually are, merely want true beliefs about the answers to questions; as Williams (2002: p. 21) emphasises, the states of nature involved in genealogical accounts are 'simplified, imaginary' environments. Furthermore, in stating the shelf life objection, Craig speaks of himself as wanting to *know* what the time is. As far as I can see, then, there is reason to think, and no reason to deny, that Craig would accept that ordinary human beings often want to know things.

in question, knowledge is no more valuable than true belief. Situations in which true belief lacks a shelf are alleged to be such a situation with respect to the stability theory. Next, one claims that, even in this type of situation, ordinary enquirers desire knowledge and not just true belief. The reasoning described in the previous paragraph can then be used to support the claim that, even in this type of situation, knowledge is more valuable than true belief. Thus, the objection concludes, the theory of the value of knowledge under scrutiny doesn't provide a full explanation of why knowledge is more valuable than true belief, because the theory doesn't explain why knowledge is more valuable in all situations in which enquirers desire knowledge and not just true belief. If the reasoning I have presented—from desiring, through preferring, to valuing and value—is correct, then a theory purporting to explain the special value of knowledge will succeed only if it meets the following adequacy condition: it must identify respects in which knowledge is more valuable than true belief that give enquirers adequate reason for preferring knowledge to true belief across the full range of circumstances in which they want to know. Indeed, it must give enquirers such reasons not just in relation to knowledge and true belief, but also in relation to knowledge and any states intermediate between true belief and knowledge. This condition is a crucial adequacy condition for philosophical theories of the value of knowledge.

I can now give an initial formulation of my argument for MKF: any theory of the value of knowledge that takes the special value of knowledge to be purely instrumental value relative to true belief will fail to meet this adequacy condition. That is, it will fail to identify respects in which knowledge is more valuable than true belief and other intermediate states that give enquirers adequate reason for preferring knowledge to such states across the full range of circumstances in which they want to know. To make this claim plausible, I will show how the argument applies to Williamson's version of the stability theory. Once this has been made clear, it will be seen that the argument plausibly generalises to any theory of the value of knowledge upon which the special value of knowledge is purely instrumental value relative to true belief.

This brings us back to the shelf life objection. We have now seen reason to think that knowledge is more valuable than true belief even in those cases in which there is allegedly no value in the persistence of true belief. What we now need to consider is whether Craig is right that 'many' true beliefs lack a shelf life in this sense. A complication here is that the stability theory doesn't commit one to any particular account of the value of true belief. Its proponent might hold that true belief is merely instrumentally valuable, (e.g.) in virtue of its action guiding role. Alternatively, they might hold that true belief is intrinsically valuable. (Plato, incidentally, seems to hold that true belief is valuable in both ways: in the *Meno*, Socrates says, 'As long as they stay put, true beliefs…constitute a thing of beauty and do nothing but good' (97e–98a).) In evaluating the shelf life objection, we will therefore need to consider the extent to which the objection depends on further assumptions about the value of true belief, assumptions one isn't committed to merely in virtue of adopting the stability theory.

This point comes out clearly in Craig's presentation of the shelf life objection. As I explained earlier, Craig's reason for thinking that there are circumstances in which the stability of a true belief makes it no more valuable is that there are circumstances in which the persistence of a true belief has no practical value for the agent. This argument obviously assumes that true belief is of no intrinsic value, and that its instrumental

value is exhausted by its action guiding role. As we have just seen, however, neither of these assumptions is a commitment of the stability theory. As things stand, then, it is open to the supporter of the stability theory to respond to Craig by rejecting either (or both) of these assumptions: that is, to hold either (a) that the instrumental value of a true belief isn't exhausted by its action guiding role, or (b) that true belief is intrinsically valuable.

The first option isn't promising. When it comes to identifying ways in which true beliefs may be instrumentally valuable besides guiding action, an obvious suggestion is that true beliefs may also be instrumentally valuable in virtue of causing the subject to form further true beliefs, (e.g.) by making conscious inferences from her existing true beliefs. This suggestion raises the question of why these further true beliefs are valuable. An immediate response is that these further true beliefs are valuable because they will, in turn, cause the subject to form even further true beliefs. Clearly, however, the fact that true beliefs have a tendency to perpetuate themselves cannot contribute towards an explanation of why true beliefs are valuable unless at least some true beliefs have value that doesn't merely consist in their tendency to cause further true beliefs. Here there are two options: one might hold either that at least some true beliefs are intrinsically valuable, or that their instrumental value isn't exhausted by their tendency to cause further true beliefs. The first option anticipates option (b) above, which I will discuss in more detail in a moment. The difficulty with the second option is that, as far as I can see, the only obvious way in which true beliefs are instrumentally valuable, besides their tendency to cause further true beliefs, lies in their action-guiding potential. This returns us to the idea that the instrumental value of true belief is ultimately instrumental value relative to successful action, with the minor addendum that a true belief may be valuable in this way either *directly*, in virtue of itself guiding actions of the agent, or *indirectly*, in virtue of causing further true beliefs which guide actions of the agent. Now, in presenting the shelf life objection, Craig doesn't anticipate the point that true beliefs can be indirectly instrumentally valuable to successful action. However, as far as I can see, this point does little to mitigate the force of his objection. Craig would presumably argue that many true beliefs will be neither directly nor indirectly instrumentally valuable relative to successful action beyond the immediate future, and it seems to me that he would be correct to do so. After all, to return to Craig's example, what further beliefs that will guide my actions will my belief about the present time cause me to form, should it persist?

These considerations suggest that nothing less than commitment to (b), the claim that true belief is intrinsically valuable, suffices in conjunction with the stability theory to explain the scope of our ordinary interest in knowledge. As the knowledge first theorist would be the first to point out, the motivation for basing an account of the value of knowledge on the claim that true belief is intrinsically valuable is unclear: if we are going to countenance an intrinsic value claim, why stop short of saying that it is knowledge rather than true belief that is intrinsically valuable, or, at least, that knowledge is *more* intrinsically valuable than true belief? But even if we let this pass, and grant that true belief is no less intrinsically valuable than knowledge, it remains doubtful that considerations of stability provide us with adequate reason for preferring knowledge to true belief.

If true belief is intrinsically valuable, regardless of its content, then there will always be value in the persistence of true beliefs, independently of the presence of any further goods relative to which true belief is held to be instrumentally valuable. The stability theorist is therefore able to say something in response to the shelf life objection. However, the shelf life objection is only one way of developing the concern that the stability theory is unable to account for the scope of our ordinary interest in knowledge. Others remain. As we saw earlier, according to Williamson's version of the stability theory, one's beliefs are more stable when one knows because one's beliefs are less likely to be rationally undermined by future evidence that comes into one's possession. It follows that in circumstances in which evidence that could rationally undermine the mere true belief that p is very unlikely or certain not to come into one's possession (call such circumstances cases of 'evidential fair weather'), one's belief that p would be at most negligibly more stable if one knew that p than if one merely truly believed p.[27] If, in these circumstances, ordinary enquirers would still want to know the answers to questions, then we appear to have reason to think that knowledge is non-negligibly[28] more valuable than true belief in cases of evidential fair weather, despite the fact 'known beliefs' are at most negligibly more stable in such cases. Thus, even if we grant that true belief is intrinsically valuable, it remains doubtful that the stability theory can fully explain the special value of knowledge.

Although I think that considerations about shelf life and evidential fair weather ultimately show that Williamson's stability theory cannot explain the scope of our ordinary interest in knowledge, there is a respect in which the argument up to this point is too simple. The argument is too simple because it ignores the fact that an enquirer may not *know* whether a true belief would lack a shelf life, or whether she is in circumstances of evidential fair weather. If she doesn't know these things, then the reasoning from desiring, through preferring, to valuing and value cannot be used to support the claim that, even in these circumstances, knowledge is more valuable than true belief. In particular, if she doesn't know these things, then it cannot be inferred from the fact that she wants knowledge and not just true belief in these circumstances that she *prefers* even 'fair weather knowledge without a shelf life' to true belief. If we aren't justified in attributing this preference to enquirers, then we cannot reason our way to the further claim that enquirers value 'fair weather knowledge without a shelf life' more highly than true belief, and thus attain evidence that 'fair weather knowledge without a shelf life' is more valuable than true belief. For this reason, the versions of the shelf life objection and the problem of evidential fair weather presented by Craig (1990) and Hyman (2010) are flat footed.

This objection highlights the fact that we cannot move straightforwardly—or, at least, as straightforwardly as has been suggested so far—from the observation that ordinary enquirers want to know in certain circumstances to the conclusion that knowledge is more valuable than true belief (and other intermediate states) in those circumstances. If there is a chance from an enquirer's perspective that a

---

[27] This point is made by Hyman (2010: p. 406).

[28] I will come back to why this might be taken to show that knowledge is non-negligibly more valuable than true belief shortly.

true belief about the answer to her question would have a shelf life, or that she is not in circumstances of evidential fair weather, then she still has some reason for preferring knowledge to true belief if the stability theory is true, even if 'fair weather knowledge with a shelf life' is no more valuable than true belief. However, even if enquirers always have *some* reason for preferring knowledge to true belief (and other intermediate states) if the stability theory is true, it is a further question whether this reason is *sufficiently strong* to justify them in having a general preference for knowledge over true belief. It is this latter point that the proponent of Williamson's stability theory must insist on if her account of the value of knowledge is to provide an adequate explanation of why we desire knowledge rather than the alternatives, and thus give a satisfactory account of the special value of knowledge.

Here we arrive at the crux of the problem for Williamson's stability theory. The reason his stability theory gives us for preferring knowledge to true belief and other intermediate states appears to be a relatively minor one: when one knows one's true belief is more stable, because it is less likely to be rationally undermined by future evidence that comes into one's possession. The strength of this reason for preferring knowledge depends on several further factors, including (i) how much value there is in the persistence of true belief, (ii) how likely one is to encounter future evidence that could rationally undermine one's belief, and (iii) the extent to which the processes through which one forms and revises one's beliefs are rational and sensitive to evidence (for further discussion of this, see footnote 17). Against this reason for preferring knowledge to true belief, one also has to consider the extent to which acquiring knowledge would be more difficult than acquiring true beliefs—if arriving at knowledge would be significantly harder, then seeking knowledge rather than true belief plausibly won't be justified by minor gains in the stability of one's beliefs. (It is the allegedly greater difficulty of acquiring knowledge that motivates the claim that, if it is to be worth pursuing, knowledge must be *non-negligibly* more valuable than the alternatives.) These points don't conclusively demonstrate that Williamson's stability theory fails to provide us with adequate reason for preferring knowledge to true belief, and thus that it cannot provide a full account of the special value of knowledge. However, I think that they do suffice to show that the claim that it does is implausible, and to shift the burden of proof firmly onto the proponent of the theory.[29,30]

This problem for Williamson's stability theory plainly has wider ramifications. If we desire knowledge rather than true belief and other intermediate states because knowledge is more valuable than these states, then any theory of the value of knowledge can potentially be queried on the grounds that, even if it succeeds in identifying a respect in which knowledge is more valuable than true belief and other intermedi-

---

[29] Note that this difficulty is exacerbated by the existence of the secondary value problem—it needs to be shown that, despite the greater difficult of acquiring knowledge, the distinctive ways in which it promotes true belief make it preferable not just to true belief *simpliciter* but to all states intermediate between true belief and knowledge.

[30] It is of course consistent with my argument that Williamson's theory *partially* explains why knowledge is more valuable than true belief. Perhaps, for example, the greater practical value of knowledge compared to true belief is merely due to the greater stability of one's beliefs when one knows. My argument is merely that Williamson's theory doesn't *fully* explain why knowledge is more valuable than true belief and states intermediate between true belief and knowledge.

ate states, that respect gives us insufficient reason for preferring knowledge to these sub-epistemic states. In particular, I think that this type of consideration is likely to undermine any theory of the value of knowledge according to which the special value of knowledge is merely instrumental value relative to true belief. Anyone wishing to endorse this claim faces two challenges. The first is to identify ways besides stability in which knowledge has greater instrumental value relative to true belief than true belief *simpliciter* and than states intermediate between true belief and knowledge. As far as I know, no such way has been proposed. The second challenge is to show that the fact that knowledge better promotes true belief in this way gives us sufficient reason, perhaps in conjunction with considerations to do with the greater stability of one's beliefs when one knows, for preferring knowledge to true belief-involving states that are insufficient for knowledge across the full range of circumstances in which we want to know. Meeting this challenge will involve engaging with the kinds of issues identified in the previous paragraph, especially those to do with the greater difficulty of acquiring knowledge. Although these considerations do not amount to a conclusive demonstration of MKF, they surely amount to a powerful argument in its favour.

This argument for MKF relies on the assumption that observations about the circumstances in which ordinary enquirers want to know are a source of evidence about the special value of knowledge, because the desire for knowledge indicates a preference for knowledge over true belief and other intermediate states, which in turn suggests that enquirers value knowledge over these alternatives. So far, this reasoning has gone unchallenged. Given its centrality to the present argument, however, more should be said. In the final section I will briefly consider the implications of rejecting the reasoning.

5 A dilemma

The assumption that ordinary enquirers desire knowledge and not just true belief or other intermediate states because knowledge is more valuable than such states is widespread in the literature on the value of knowledge. One piece of evidence for this is that philosophers formulate the Meno problem and the secondary value problem interchangeably as problems about our desires,[31] problems about our preferences,[32] and as problems about value.[33] In making this assumption, philosophers take for granted something like the reasoning outlined in the previous section, from desiring, through preferring, to valuing and value. However, though natural, this reasoning

---

[31] For example, Pritchard and Turri (2017) offer the following formulation of the secondary value problem: 'Whereas the *Meno* problem concerns the question of why knowledge is more valuable than mere true belief, the secondary value problem concerns the issue of why knowledge is more valuable than any proper subset of its parts. That is, why do we specifically *desire* [my emphasis] knowledge rather than any epistemic standing that falls short of knowing (including, but not restricted to, mere true belief)?'.

[32] For example, the *Internet Encyclopedia of Philosophy* formulates the Meno (or primary value) problem as follows: 'Knowledge and true belief both tend to be things we want to have, but all else being equal, we tend to *prefer* [my emphasis] to have knowledge over mere true belief. The Primary Value Problem is the problem of explaining why that should be the case' (Bondy 2019).

[33] For a particular clear instance of the tendency to conflate these different kinds of question, see the first paragraph of Sect. 1 of Jones (1997).

certainly isn't above suspicion. Given the way it can be exploited to challenge accounts of the value of knowledge, some may be tempted to reject it. In this section I will briefly discuss the implications of doing so for proponents of philosophical theories of the value of knowledge.

The central point I want to make is this: if proponents of theories of the value of knowledge choose to reject the reasoning, then they are in serious danger of cutting the ground from beneath their feet. The reason for this is that it isn't clear that there is any 'intuition' that knowledge is more valuable than true belief, and than states intermediate between true belief and knowledge, independently of the reasoning from desiring, through preferring, to valuing and value. One way of bringing this out is to imagine how we would feel if we had an explanation of why enquirers desire knowledge, rather than true belief and other intermediate states, that didn't appeal to the idea that knowledge is a specially valuable state. If someone tried to impress on us that our theory wasn't enough, that we still needed to explain why knowledge is a specially valuable state, I think our reaction would be one of scepticism. 'Look', we would say, 'we already have an explanation of why we want to know, and not just to be in these other states. Our everyday concern with knowledge is already intelligible to us. Why insist on the further idea that knowledge must also be a specially valuable state?'. If this is right, then the reasoning from desiring, through preferring, to valuing and value isn't merely a further consideration shaping philosophical conceptions of the explananda for theories of the value of knowledge: it is the foundation on which the Meno problem and the secondary value problem rest. And, if this is right, then rejecting this reasoning has an embarrassing consequence for proponents of philosophical theories of the value of knowledge. To be sure, it doesn't follow from the incorrectness of the reasoning that their explanations of why knowledge is a specially valuable state are incorrect. However, it does follow that we have no reason to think that knowledge is a specially valuable state *in advance* of a theoretical account of why it is. If we have no such reason, then it is misleading to present theories of the value of knowledge as grand solutions to deep conundrums. They are mere philosophical curiosities.

An upshot of the argument of this paper, then, is that proponents of philosophical theories of the value of knowledge face a dilemma. On the one hand, they might accept the reasoning from desiring, through preferring, to valuing and value, and so accept the assumption that we desire knowledge and not just true belief or other intermediate states because knowledge is a specially valuable state. In doing so, they avail themselves of evidence that knowledge is a specially valuable state that is independent of any theoretical account of why knowledge is a specially valuable state. However, they also risk exposing themselves to the objection that their account of the special value of knowledge fails to identify adequate reasons for ordinary enquirers to prefer knowledge to true belief and other intermediate states. In the previous section, I argued on these grounds that Williamson's stability theory fails to provide a full account of the special value of knowledge, and that this result plausibly generalises to any theory of the value of knowledge according to which the special value of knowledge is merely instrumental value relative to true belief. Alternatively, proponents of philosophical theories of the value of knowledge might reject the reasoning from desiring, through preferring, to valuing and value. Consequently, they will no longer be dogged by concerns that their theory fails to supply ordinary enquirers with adequate reasons for

preferring knowledge to the alternatives. However, they will also have deprived their theory of its *raison d'être*.

Given the foregoing considerations, two issues in need of further investigation are the following. The first is the extent to which other accounts of the value of knowledge are undermined by objections based on the need to explain why ordinary enquirers want to know. In this paper I have made a case for the minimal knowledge first value claim by deploying an instance of this objection against the stability theory, but I suspect that the objection applies more widely. Ascertaining just how widely it applies is necessary for gauging the sharpness of the dilemma here for proponents of philosophical theories of the value of knowledge. The second issue requiring further investigation is whether we should accept the reasoning from desiring, through preferring, to valuing and value. Although this reasoning appears foundational to philosophical debates about the value of knowledge, it has not been subjected to careful scrutiny. Such scrutiny must be given; though natural, the reasoning is far from obviously correct.[34]

References

Bondy, P. (2019). *Epistemic normativity*. The Internet Encyclopedia of Philosophy. ISSN 2161-0002. https://www.iep.utm.edu/home/about/. February 6, 2019.

Cassam, Q. (2009). Can the concept of knowledge be analysed? In P. Greenough & D. Pritchard (Eds.), *Williamson on knowledge* (pp. 12–30). Oxford: Oxford University Press.

Cassam, Q. (2014). *Self-knowledge for humans*. Oxford: Oxford University Press.

Craig, E. (1990). *Knowledge and the state of nature*. Oxford: Oxford University Press.

Dutant, J. (2014). The value and normative role of knowledge. In J. Dutant, D. Fassio, A. Meylan (Eds.) *Liber Amicorum Pascal Engel* (pp. 200–227). University of Geneva. http://www.unige.ch/lettres/philo/publications/engel/liberamicorum. ISBN 978-2-8399-1562-5. February 6, 2019.

Fricker, M. (2009). The value of knowledge and the test of time. *Royal Institute of Philosophy Supplement, 64,* 121–138.

Gibbon, J. (2013). *The norm of belief*. Oxford: Oxford University Press.

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Hawthorne, J., & Stanley, J. (2008). Knowledge and action. *Journal of Philosophy, 105*(10), 571–590.

Hyman, J. (2010). The road to Larissa. *Ratio, 23*(4), 393–414.

Jones, W. E. (1997). Why do we value knowledge? *American Philosophical Quarterly, 34*(4), 423–439.

Kripke, S. (2011). On two paradoxes of knowledge. In S. Kripke, *Philosophical Troubles* (pp. 27–51). Oxford: Oxford University Press.

Kvanvig, J. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge: Cambridge University Press.

---

[34] For critical discussion of this reasoning, see my 'Enquiry and the Value of Knowledge' (2019).

Millar, A. (2011). Why knowledge matters. *Proceedings of the Aristotelian Society Supplementary, 85,* 63–81.

Olsson, E. J. (2007). Reliabilism, stability, and the value of knowledge. *American Philosophical Quarterly, 44,* 343–355.

Plato (2005) *Meno and other dialogues* (R. Waterfield, Trans.). Oxford: Oxford University Press.

Pritchard, D. (2007). Recent work on epistemic value. *American Philosophical Quarterly, 44*(2), 85–110.

Pritchard, D. (2010). Knowledge and understanding. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *The nature and value of knowledge: Three investigations* (pp. 3–90). Oxford: Oxford University Press.

Pritchard, D., & Turri, J. (2017), The value of knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*. https://plato.stanford.edu/archives/win2017/entries/knowledge-value/. February 6, 2019.

Sutton, J. (2007). *Without justification*. London: MIT Press.

Williams, B. (2002). *Truth and truthfulness: An essay in genealogy*. Princeton: Princeton University Press.

Williams, B. (2005). *Descartes: The project of pure enquiry*. London: Routledge.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

Williamson, T. (2018). Acting on knowledge. In J. A. Carter, E. C. Gordon, & B. Jarvis (Eds.), *Knowledge first: Approaches in epistemology and mind* (pp. 163–181). Oxford: Oxford University Press.

Walker, B. (2019). Enquiry and the value of knowledge. Unpublished.