CrossMark

The internalist virtue theory of knowledge

Ralph Wedgwood[1] ⬤

Abstract Here is a definition of knowledge: for you to know a proposition p is for you to have an outright belief in p that is correct precisely because it manifests the virtue of rationality. This definition resembles Ernest Sosa's "virtue theory", except that (a) on this definition, the only virtue that must be manifested (at least to some degree) in all instances of knowledge is rationality, and (b) no reductive account of rationality is attempted—rationality is assumed to be an irreducibly normative notion. This definition is compatible with "internalism" about rationality, and with a form of "pragmatic encroachment" on the conditions of rational outright belief. An interpretation is given of this definition, and especially of the sense of 'because' that it involves. On this interpretation, this definition entails that both safety and adherence are necessary conditions on knowledge; it supports a kind of contextualism about terms like 'knowledge'; and it provides resources to defend safety, adherence, and contextualism, against some recent objections.

Keywords Ernest Sosa · Knowledge · Virtue · Rationality · Safety · Adherence · Contextualism

1 "The aim of belief" revisited

Many philosophers have claimed that *beliefs aim at the truth*. We can raise many questions about how to understand this claim. For example, Sosa (2015, p. 24) has argued that it is literally true that beliefs aim at the truth: at the level of what Sosa calls

✉ Ralph Wedgwood
wedgwood@usc.edu

1 School of Philosophy, University of Southern California, 3709 Trousdale Parkway, Los Angeles, CA 90089-0451, USA

"functional" belief, he suggests that this aim is "teleological, like that of perception"—while at the level of what he calls "judgmental" belief, he suggests that "the aim … is like that of intentional action' (25). I have myself also defended the claim that belief aims at the truth, but only if the claim is understood as a metaphorical way of conveying an essentially *normative* point—the point that whenever someone has a belief, that belief is *correct* if and only if the proposition believed is true (Wedgwood 2002).

In this discussion, however, I shall not worry about how best to interpret the use of the term 'aim' in this claim. I shall simply assume that some reasonable interpretation of this occurrence of the term can be found. Moreover, to keep things simple, I shall here ignore partial beliefs or levels of credence or confidence, and restrict my attention here to *full* or *outright* beliefs.[1] So, I shall focus on the following version of this claim: whenever you rationally have a full or outright belief in a proposition p, your aim is to believe p if and only if p is true. As I shall put it, when you rationally believe p, your aim is to have a belief that is *correct*, or *gets things right*, about p.

Among those who accept that in some sense belief aims at the truth, there are those who also suggest that in some corresponding sense, the *means* that rational thinkers use to pursue this aim is thinking *rationally*. As Pettit (1993, p. 68) put it: "Thinking is the intentional attempt to submit oneself, ultimately to the regimen of truth, proximally to the discipline of rationality."[2] The suggestion that I shall explore here, then, is that when one rationally believes p, one's rationally believing p is the means that one uses to pursue the aim of having a correct belief about p.

In what sense, exactly, is one's rationally believing p one's "means" to having a correct belief about p? I shall answer this question in Sect. 2 below, by clarifying what exactly I mean by describing rational believing as one's "means" toward correct believing. Before turning to that question, however, I want to sketch the main proposal of this paper—the proposal that we can use this idea of rational believing as the "means" that one uses to pursue correct believing in order to give an account of *knowledge*.

Whenever one uses means to pursue an end, it is possible for one's end to be realized, but not *because* of one's use of these means. For example, suppose that my aim is that you should be dead by the end of the day today, and the means that I use to achieve this end is by slipping a lethal poison into your morning coffee. It could happen that by a strange fluke you have eaten something that acts as an antidote to this poison, but die anyway in a car accident on the way to work. In this case, my aim is realized, but not because of the means that I used to pursue the end.

Even a *causal* connection between the means that I use to pursue the end and the realization of the aim is not enough to make it true that I *succeed* in *achieving* the aim. This point can be illustrated by one of the classic examples that are due to Donald Davidson (1980, p. 78):

> A man may try to kill someone by shooting at him. Suppose the killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death.

---

[1] For an account of the nature of full or outright beliefs, see Wedgwood (2012a).

[2] I have made this suggestion in my own earlier work; see Wedgwood (2002, 2013).

Here, there is a causal chain between the means that the agent employs and the realization of his aim. But the causal chain is *deviant*: the way in which the agent's employment of the means causes the realization of the aim does not count as the agent's succeeding in achieving the aim. For the agent to succeed in achieving the aim, there must be the *right kind* of explanatory connection between the agent's employment of the means and the realization of the aim.

Both of these two kinds of case have analogues involving belief. You might rationally believe *p*, and *p* might be true, but it could be a lucky fluke that both these conditions hold. As we shall see, this is true in each of the cases that were made famous by Gettier ([1963](#)). Moreover, even a causal connection between your rationally believing *p* and your correctly believing *p* seems not to be enough to ensure that there is the right kind of explanatory connection. Perhaps there are two demons working against each other in your environment: an evil demon who gives you hallucinatory experiences, and a second benevolent demon who changes your environment in such a way as to ensure that whenever you rationally believe a contingent proposition *p*, the proposition *p* is true. In this case, the fact that you rationally believe *p* causes the second demon to make it the case that *p* is true. But this is still a deviant causal chain—not the right kind of explanatory connection that is needed to rule out such deviant causal chains.

It is a famous problem how exactly to characterize the kind of explanatory connection that is needed to rule out such deviant causal chains. One promising approach is to characterize the required explanatory connection in terms of the manifestation of an appropriate *disposition*.[3] Specifically, there are certain dispositions that are presupposed by everyday folk psychology as in a sense "basic dispositions" of the agents whose actions and attitudes folk psychology is equipped to explain. Some of these are dispositions to be such that, in appropriately *normal* cases in which the agent employs means of the relevant kind, there is at least a reasonable chance of the relevant aim's being realized. Then the explanatory connection between the employment of the means and the realization of the aim is of the appropriate non-deviant kind just in case it consists in the *manifestation* of dispositions of this kind.

I shall not inquire here whether this approach provides the right solution to the problem of deviant causal chains in general. Instead, I shall simply try to characterize a certain kind of explanatory connection that can hold between (i) your rationally believing *p* and (ii) your correctly believing *p*. The proposal that I shall explore here is that this kind of explanatory connection is what we need to give an account of *knowledge*.

More specifically, according to this proposal, if you have an outright belief in *p*, this counts as a case of *knowing p* if and only if it is a case of believing *correctly* precisely *because* it is case of believing *rationally* (where the phrase 'precisely because' is interpreted as indicating an explanatory connection of the kind that I shall characterize). As I put it in some of my earlier work (Wedgwood [2002](#), p. 283): "one knows something when one has, through one's own rational efforts, succeeded in achieving the aim of

---

[3] For this approach to characterizing this sort of explanatory connection, see Wedgwood ([2006](#)).

believing the truth". In the rest of this discussion, I shall try to explain and defend this proposal about the nature of knowledge.

2 Rationality as an (internalist) virtue

As I noted above, we need to clarify in what sense it is being proposed that rationally believing p is the "means" that the thinker uses to pursue the "aim" of having a correct belief about p. To do this, it will be helpful to clarify exactly what assumptions about rationality we will be relying on. The goal of this section is to explain these assumptions about rationality. This goal does not require giving a complete account of rationality—on the contrary, the proposal of this paper is designed to be compatible with a wide range of accounts of rationality. But certain assumptions about rationality will be important for the proposal that follows. The most important of these assumptions is that rationality is in a sense a *virtue*.

There is a familiar distinction in epistemology between the propositions that an agent has *propositional justification* for believing, and the beliefs that the agent holds in a *doxastically justified* manner. An agent might believe a proposition p at the same time as p's being a proposition that the agent has propositional justification for believing, but it might be a lucky fluke that both these two conditions hold at the same time. In this case, the agent would not count as believing p in a doxastically justified manner.[4] Exactly the same distinction can be drawn in terms of rationality. An agent might believe a proposition p at the same time as p's being a proposition that *it is rational* for her to believe—even if it is a lucky fluke that both these two conditions hold at the same time. In this case, the agent would not count as *rationally believing p*.

As I have argued elsewhere (Wedgwood 2017, pp. 140–142), this seems to be precisely analogous to a distinction that Aristotle drew in the *Nicomachean Ethics* (1105a17–b9). An agent might perform an act of type A at the same time as being in a situation in which *it is just* for her to perform an act of type A—even if it is a lucky fluke that both these two conditions hold at the same time. In this case, the agent might be doing a *just act*, but she would not count as *acting justly*.

To fix ideas, let us suppose that every "act" involves an agent, an act-type, and a time. If an act is a case of the agent's acting justly, the following two conditions must hold. First, the act must be a just act—it must involve an agent, an act-type, and a time such that it is just for the agent to perform an act of that type at that time. Secondly, the agent must be manifesting an appropriate disposition—specifically, a disposition that non-accidentally tends to result in the agent's doing just acts. It seems that having this disposition is at least part of what is involved in possessing (to at least some degree) the virtue of justice, or being a just person.

In a similar way, let us suppose that every "belief" involves an agent, the property of believing a certain proposition p, and a time. Then, I propose, if a belief is a case of the agent's rationally believing p, the following two conditions must hold. First, p must be a proposition that it is rational for the agent to believe at the relevant time. Secondly, the agent must be manifesting an appropriate disposition—specifically, a

---

[4] For an interesting contemporary discussion of this distinction within epistemology, see Turri (2010).

disposition that non-accidentally tends to result, at each time when the agent manifests the disposition, in her believing a proposition that it is rational for her to believe at that time. I shall refer to dispositions of this kind as *rational dispositions*.[5]

In this way, we can identify three good features associated with rationality. *Being a belief in a proposition that it is rational for the agent to believe at the time* is one good feature, exemplified by beliefs; *having a rational disposition* is a second good feature, exemplified by agents; and a belief's being a case of *the agent's believing rationally* is a third good feature, exemplified by a belief whenever the belief has the first good feature, and the agent's believing the proposition in question is the manifestation of a disposition that constitutes the agent's having the second good feature.

There is a parallel trio of good features associated with justice. Being a just act is a good feature of acts; having a disposition that non-accidentally tends to result in one's performing such just acts is a second good feature, exemplified by agents; and an act's being a case of the agent's acting justly is a third good feature, exemplified by an act whenever the act has the first good feature, and the agent's performing the act is the manifestation of a disposition that constitutes her having the second good feature.

I shall suppose here that we can define the third good feature—namely, rationally believing a proposition p—in terms of the first two good features—being a belief in a proposition that it is rational for one to believe at the time, and having a rational disposition. But I shall not take any stand here on whether either of the notions of the first two good features is more fundamental than the other. (Similarly, I shall not take any stand on whether the notion of a just act is more fundamental than the notion of the dispositions that are characteristic of the just person.) Perhaps the two notions are coeval—interrelated notions neither of which is more fundamental than the other. All that I am assuming here is that these notions are related in the way that I have described.

Because of these parallels between rationality and a paradigmatic virtue like justice, it seems reasonable to use the term 'virtue' so that rationality itself is a kind of virtue, or a collection of virtues. Specifically, I shall say that each of the rational dispositions that I have described is itself a "rational virtue". Manifesting these dispositions is manifesting (to at least some degree) these rational virtues.

Beyond this fundamental conception of rationality as a kind of virtue, no particular account of rationality will be assumed here. So, in particular, in this discussion I shall remain completely neutral on the debate between internalists and externalists about rationality. As it happens, in fact, the arguments for internalism about rationality seem to me to be entirely compelling. So the proposal about knowledge that I shall develop in the rest of this paper is designed to be compatible with an internalist theory of

---

[5] An anonymous reviewer for this journal raised the issue of a belief that results from the operation of *two* different dispositions—one of which is rational, while the other is not. The question here is whether (a) both these dispositions are causally necessary parts of a single sufficient causal explanation, or (b) each of these dispositions is sufficient by itself (so that we have a case of causal overdetermination). I am inclined to think that in the former case, the belief is *not* a case of believing rationally, but in the latter case, it *is* a case of believing rationally. So, more precisely, what is required for believing rationally is that the agent must manifest rational dispositions, and the manifestation of these rational dispositions must be a sufficient explanation of the agent's having the belief. But it is not required that this should be the *only* sufficient explanation of the belief.

rationality. But strictly speaking I shall not need to assume an internalist view of rationality here.[6]

In general, this proposal about knowledge is designed to be compatible with a wide range of different conceptions of rational belief. For example, it is compatible with both foundationalist and coherentist conceptions of rational belief.[7] It need take no stand on whether rationality is primarily exemplified by enduring belief-states or by processes of belief-revision, or on whether the rationality of a belief-state at an instant in time t supervenes purely on how things are at that instant t, or on how things are not just at t but also over some immediately preceding period. It is also compatible with the idea that has come in recent years to be known as "pragmatic encroachment"—that is, with the idea that *pragmatic factors* (such as the needs, interests, and values at stake in the agent's situation) may make a difference to whether a proposition p counts as one that it is rational for the agent to believe.

If this is what I mean by talking about an agent's "rationally believing" a proposition p, what might it mean to suggest that rationally believing p is the "means" that one uses to pursue the aim of getting things right about p? I propose the following interpretation: when one rationally believes p, one's believing p is the manifestation of these rational dispositions; and as I shall explain, normally—under favourable circumstances—the fact that one's believing p is the manifestation of such rational dispositions provides a kind of *explanation* of why one gets things right about p.

According to the central proposal that I shall defend here, when the fact that one's believing p is the manifestation of such rational dispositions in an appropriate non-deviant way explains why one has a correct belief about p, this counts as one's *knowing p*. Given my conception of rationality as a kind of virtue, this proposal is equivalent to the thesis that if you have an outright belief in a proposition p, this counts as a case of your knowing p if and only if it is a case of your believing *correctly* precisely *because* it is a case of your *manifesting rational virtues*. In this way, this proposal is closely akin to the "virtue epistemology" of Sosa (2007, Lecture 2).

However, there are two crucial differences between the version of virtue epistemology that I am proposing here and Sosa's version. First, according to what I am proposing, the kind of virtue that must be exemplified in cases of knowledge consists of what I am calling *rational virtues*. No other cognitive virtues or skills of any kind need be exemplified—whereas these rational virtues must be exemplified, at least to some degree, in *every* case of knowledge.

Secondly, I make no attempt to analyse these rational virtues. In particular, I shall not follow Sosa (2007, p. 29) in attempting to analyse these virtues in the "reliabilist" style, in terms of a reliable disposition to believe truths of a certain kind, or the like; instead, I shall treat the concept of "rationality", as it appears here, as an *irreducibly normative* or *evaluative* notion.

These two differences between my proposal and Sosa's approach are, at least arguably, improvements. On the first point, as Lackey (2007) has insisted, beliefs

---

[6] For my most recent statement of my arguments for internalism about rationality, see Wedgwood (2017, Chap. 7).

[7] It is also compatible with the mixed view that I have defended—"Coherentism for enduring beliefs, foundationalism for mental events [of belief formation or belief revision]"; see Wedgwood (2012b, p. 286).

based on casual testimony can count as knowledge, even though they need not stem from any special expertise or skill of the believer. But plausibly, no belief counts as knowledge unless it manifests rational virtues, at least to a sufficient degree. Even a belief based on casual testimony only counts as knowledge if the believer is being sufficiently rational in trusting the testimony in question. Of course, as the cases of Gettier (1963) have in effect taught us, even if a true proposition is rationally believed, this is not *sufficient* for the belief to count as knowledge. Nonetheless, it is plausibly a *necessary* condition of a belief's counting as knowledge that the relevant proposition must be both correctly and (to a sufficient degree) rationally believed.

On the second point, merely manifesting a reliable disposition to believe truths seems not to be sufficient either for rationality or for knowledge. Many philosophers hold that this point is illustrated by Bonjour's (1980) "clairvoyance" cases: according to these philosophers, in these cases the belief in question is a result of a reliable disposition, but fails to count as knowledge because it is insufficiently rational. As I have emphasized, I am not defending any particular account of rationality here. But if it is true that the beliefs in these BonJour cases fail to count as knowledge because they are insufficiently rational, there will presumably be some account of rationality that gives the correct explanation of why these beliefs are insufficiently rational. Whatever this account turns out to be, this account can simply be plugged into the view of knowledge that I am proposing here. If it is true that the BonJour cases fail to count as knowledge because they are insufficiently rational, then leaving an open space in our theory of knowledge—to be plugged by the correct account of rationality, whatever it turns out to be—seems the best way of ensuring that our theory is equipped to deal with these cases.

3 Why knowledge requires "safety"

According to the proposal that I am exploring here, if you have an outright belief in a proposition *p*, this counts as a case of your *knowing p* if and only if it is a case of your believing *correctly* precisely *because* it is a case of your believing *rationally* (that is, a case of your manifesting rational virtues to a sufficient degree).

The crucial task for me, in exploring this proposal, is to give a characterization of this kind of 'because'. There seem to be several different kinds of 'because', each of which expresses a different explanatory connection. To evaluate this proposal about the nature of knowledge, we need to have a better understanding of what sort of explanatory connection it involves.

The crucial point, I believe, is that every explanation reveals the particular case that is under consideration as an instance of a *more general pattern*—where this more general pattern obtains, not only in the actual world, but also in a suitable range of close non-actual possible worlds. In other words, every explanation implies a kind of *modally robust generality*.

Let us suppose that each of the explanations that we are interested in concerns a *case*, where any such "case" is, in effect, a *centred world*—that is, in effect, a possible world with a certain state of affairs picked out as the centre of the world. For each of these cases, let us suppose that the state of affairs at the case's centre is a state of

affairs consisting in a certain *thinker*'s having a certain *doxastic attitude* towards a certain *proposition* at a certain *time*. Thus, one such case might be centred on your now having an outright belief in the proposition that there is no largest prime number; and another such case might be centred on my suspending judgment yesterday about whether Julius Caesar's horse crossed the Rubicon with its right leg first or not. As I shall put it, every case has a "target proposition", and involves the relevant thinker's having a certain doxastic attitude—some kind of broadly belief-like attitude—towards that proposition at the relevant time.

Each of the explanations that we are interested in gives an account of why one condition (the *explanandum*) holds of a certain case, by pointing to a different condition (the *explanans*), which also holds of the case. To a rough first approximation, the kind of explanation that we are interested in is the kind that seeks to identify an *explanans* that is a *sufficient condition* for the *explanandum*—that is, an *explanans* such that, in all the relevant close actual and non-actual possible worlds, any case in which the *explanans* holds is also a case in which the *explanandum* holds.

Strictly speaking, however, this is only a rough first approximation. There are at least two reasons for this. First, in many of the explanations that we give, it is common for the *explanandum* and the *explanans* to be indicated roughly, in terms of the particular case that we are interested in. Thus, the *explanandum* might be, not simply *that there was a fire*, but more specifically *that there was a fire more-or-less like the fire that occurred in this case*.

Secondly, in practice, the explanations that we accept as adequate rarely succeed in identifying a *sufficient* condition for the *explanandum*. We might accept that the fact that *there was a short circuit* explains why there was a fire. But of course, the fact that there was a short circuit by itself was not strictly sufficient for the fire: the presence of oxygen and inflammable materials were also required. Thus, in accepting that the short circuit explains the fire, we are presupposing a background of *normal conditions*. The presence of oxygen in the atmosphere and inflammable materials like curtains and furniture are presupposed as part of this background of normal conditions.[8]

Putting these points together, the hypothesis that the case of your now believing *p* is a case of your believing correctly *precisely because* it is a case of your believing rationally can be unpacked in the following way. It is, in effect, equivalent to the following hypothesis:

> All the relevant close (actual and non-actual) possible cases that are *sufficiently similar* to the case under consideration, both (a) with respect to what makes your belief *rational*, and (b) with respect to the extent to which, and the way in which, your conditions are *normal*, are also similar with respect to their involving a *correct belief*.

---

[8] A further complication with these explanations is that even given this presupposed background of normal conditions, the regularities or laws of nature that underlie the explanation need not be strictly *deterministic*. So, it is enough if—within the relevant domain of close possible worlds—any case in which background conditions are normal, and in which the *explanans* holds, is also a case in which there is a sufficiently high *chance* of the *explanandum*'s also holding. To keep things simple, however, I shall ignore this complication in this discussion.

So, this kind of 'because' effectively picks out a domain of sufficiently close sufficiently similar cases, similar with respect to the *explanans* (your thinking rationally), and with respect to the background normal conditions. For the explanatory hypothesis to be true, something sufficiently similar to the *explanandum* (your believing correctly) must be true throughout this domain of cases.[9]

What exactly is the notion of "normality" that is being deployed here? It seems that there is a certain kind of normality that we presuppose as part of the background whenever we give folk-psychological explanations of why a thinker is right about some question. It is this notion of normality that I am invoking here. It seems plausible that our grasp of this notion of "normality" is partly informed by our competence with its role as a background condition for ascriptions of knowledge.[10] In principle, however, it is a more general notion, since it would also appear in folk-psychological explanations of cases that do not involve knowledge—such as explanations of why a thinker has a high credence (falling short of outright belief) in a true proposition, or a low credence in a false proposition.

The need to introduce such a "normality" condition is particularly clear if the internalist conception of rationality is correct. According to this proposal, knowledge requires the pattern "Similar rationality—similar correctness" to hold throughout the relevant sufficiently "close" cases; but these are not just the cases that are similar with respect to the purely *internal* rationality of the thinker's thinking. These cases must also be sufficiently similar with respect to the degree to which, and the way in which, the thinker's *external* conditions are "normal". It is only throughout this range of similarly normal close cases that the cases that are similar with respect to what makes the belief in question rational must also be similar with respect to involving a correct belief.

It follows that, in every similarly normal close possible case in which you have a belief on a similar topic that is rational in a similar way, the proposition believed in that case must be *true*—since if in one of these cases, the proposition believed were false, that case would not be similar with respect to involving a correct belief. In other words, it cannot be that in any of these sufficiently similar close cases you have a false belief.

This condition is a version of what philosophers like Sosa (1999) and Williamson (2000) have called "safety". A belief *b* that meets this condition is in a clear sense *safe from error*: it could not easily happen that a belief that resembles *b* in all the relevant respects would be incorrect. In other words, it is a consequence of the proposal about knowledge that I am exploring here that knowledge requires safety. If a belief is to count as knowledge, the belief must be safe.

---

[9] The problem of identifying the relevant range of cases is akin to what Conee and Feldman (1998) have called the "generality problem" for reliabilist theories of knowledge. If my account of knowledge is combined with a substantive theory of rationality, this theory of rationality should help to identify the relevant range of cases. The theory of rationality will supply a way of specifying the dispositions that are manifested in rationally-held beliefs; and the relevant cases must be similar with respect to the dispositions of this kind that are manifested in them.

[10] Compare the claims that Williamson (2000, p. 100) about how to interpret a "reliability" condition on knowledge.

This point—that the version of the virtue theory of knowledge that I am exploring here entails that knowledge requires safety—might seem surprising. This version of the virtue theory identifies knowledge with something that is akin to what Sosa calls "apt belief"; and Sosa has argued that a belief can be "apt" without being safe.[11] As I shall explain in Sect. 6, however, in making this argument, Sosa is focusing on some *stronger* forms of "safety" than the kind of safety that I am discussing here. Sosa is right that aptness does not entail any of these stronger forms of safety, but that is compatible with the point that I have just argued for, that aptness entails the weaker kind of safety that I have been discussing here.

Even though the kind of "safety" that I am discussing here is weaker than some other kinds, it is still strong enough to rule out the famous cases of Gettier (1963). Suppose that on the basis of strong evidence, Sarah believes the proposition that *Janet will get the job, and Janet has ten coins in her pocket.* (For example, suppose that Sarah has been assured by the chair of the hiring committee that Janet will get the job, and Sarah has also carefully counted the coins in Janet's pocket just five minutes ago.) From this proposition Sarah then infers the further conclusion that *the person who will get the job has ten coins in her pocket.* As it happens, this further conclusion is true—though not because Janet will get the job. On the contrary, it is true because Sarah herself will get the job, and (as it happens) she also has ten coins in her pocket.

Sarah's circumstances in this case are mildly abnormal—she has received misleading testimony from the chairman of the hiring committee. So, we need to consider a range of close possible cases that are abnormal in a similar way and to a similar degree, in which Sarah has a belief on a similar topic that is rational in a broadly similar way. In one of these cases, she believes exactly the same propositions as in the actual case, but she in fact has nine coins in her pocket rather than ten. In this case, the proposition that she believes in this case *The person who will get the job has ten coins in her pocket* is false. So, the belief that Sarah has in the actual case is unsafe.[12]

Similarly, this kind of safety is also sufficient to rule out Carl Ginet's "barn façade" case (Goldman 1976). You are driving through a region that is full of papier-mâché barn façades, which from the road look just like real barns. As you pass these barn façades, you form a belief about each—a belief that you could express by saying 'That is a barn'. After forming four such beliefs while driving past papier-mâché barn façades, you then drive past a real barn, and in this fifth case, in exactly the same way, you form a belief that you could express by saying 'That is a barn'. In this fifth case,

---

[11] More recently, Sosa (2015) has distinguished between two kinds of belief—"functional belief" and "judgment", where a "judgment" involves an act of "affirmation" that is intentionally aimed at the truth—and has argued that "aptness of judgment *entails* safety of affirmation" (79). However, Sosa (2015, p. 80) still explicitly claims: "an affirmation can be apt without being safe, without being one that *would* be true if made". This is what I am denying here: my claim is that *no* kind of belief can be "apt" without also being in a way "safe".

[12] Christoph Kelp, the editor of this volume, has raised the following objection: "What if all of this is set up by a powerful demon who sees to it that (i) Sarah gets false testimony about who gets the job and (ii) that all propositions in the range of similar propositions that Sarah might rationally infer from it are true at the relevant nearby worlds?" But in *this* case, the agent's circumstances are wildly abnormal. So, there are other cases, which are equally normal, where the demon has a slightly different plan, and so sees to it that all (or at least most) of the propositions that Sarah might rationally infer from this false proposition are *false*. Thus, introducing manipulative demons into the case will not make it safe in the relevant way.

unlike in the earlier four cases, your belief is correct: the proposition that you believe is true. Since you know nothing about these strange papier-mâché barn façades, we may assume that all five beliefs are equally rational.

In this case, your circumstances are again somewhat abnormal—you are surrounded by papier-mâché barn façades. So, we need to consider a range of close possible cases that are abnormal in a similar way and to a similar degree, in which you have a belief on a similar topic that is rational in a broadly similar way. But it seems that the earlier four cases, in which you formed a false belief that you expressed by saying 'That is a barn', are sufficiently similar in the relevant respects. So, the belief that you have in the fifth case is unsafe. In this way, this approach can explain why these beliefs do not count as knowledge. Even though these beliefs are both rational and correct, it is too much of a fluke that rationality and correctness coincide in these cases: they are not cases of your believing correctly precisely because they are cases of your believing rationally.

At the same time, this version of safety does not have the false implication that beliefs based on causal testimony cannot be safe. For example, suppose that on arriving in an unfamiliar city, you ask a random passer-by for directions to a well-known landmark, and since the answer that you receive sounds not implausible, and is given in a confident tone of voice, you rationally believe what you are told. Now, in the social circumstances that count as normal for people like you and me, it would not easily happen that a passer-by would give a plausible-sounding answer in this confident tone about the location of a well-known landmark unless they were sincere and well informed about the matter. Suppose that in fact, your circumstances are normal in precisely this way. Then, in all similarly normal cases, in which you believe a similar proposition in a similarly rational way, the proposition believed is true. In our social world, rational beliefs based on casual testimony in this way are safe.[13]

4 Why knowledge requires "adherence"

Suppose that all the relevant close cases that are sufficiently similar to the actual case—both (a) with respect to what makes your belief *rational* and (b) with respect to what makes your conditions *normal*—are also similar with respect to their involving a *correct belief*. For any of these cases to be similar to the actual case with respect to involving a correct belief, the case's target proposition must be true, and you must *believe*, or at least have *high credence* in, the proposition. If you did not at least have a high credence in the proposition, the case would clearly not be similar with respect to involving a correct belief.[14] In general, in each of these cases, the case's target

---

[13] This is how I would answer the objection of Lackey (2008, p. 34), that any version of the virtue theory that is strong enough to rule out the Gettier cases will also incorrectly rule out beliefs that are based on casual testimony. On my version of the theory, the force of 'because' in the formula 'correct *because* rational' is exhausted by the safety and adherence conditions that I articulate here, and it seems clear that beliefs based on casual testimony, if they are rationally held, can meet both the safety and adherence conditions.

[14] Admittedly, according to some views, outright belief and high credence are not similar states at all (as an anonymous reviewer for this journal has reminded me). Elsewhere (Wedgwood 2012a), I have argued that there is a certain kind of credence—"practical credence", as I called it—such that outright belief and

proposition must be true, and you must either believe or at least have high credence in that proposition.

This entails a version of the fourth condition that Robert Nozick (1981, p. 176f.) imposed on knowledge—"adherence". The relevant version of adherence is the following. For your belief to "adhere" to the truth in this way is for it to be such that, if things were at most slightly different from how they actually are, but the case were otherwise sufficiently similar to the actual case in all the relevant respects, and the case's target proposition were true, you would still *believe* or at least *have high credence in* that proposition.

To understand this kind of "adherence", we need to know more about what it is for the case to "sufficiently similar to the actual case in all the relevant respects". I propose that for a case to be sufficiently similar in these respects, the following two conditions must hold:

   a. You have a body of evidence that (i) you might easily have had, in conditions that are similarly normal in the relevant way, and (ii) is at most only *slightly* different from the evidence that you have in the actual case—such as a body of evidence that *properly includes* everything in your actual evidence, but also a few other pieces of evidence as well.

   b. You respond to this evidence, in a similarly rational way, by having some doxastic attitude towards the case's target proposition.

Adherence *fails* to hold just in case there is at least one sufficiently similar case of this kind—in a world in which things are at most slightly different from how they actually are, and the case's target proposition is true—in which your credence in the proposition is *much lower* than in the actual case, or in which you totally *suspend judgment* about that proposition.

This version of adherence is strong enough to rule out Harman's (1968, p. 172f.) "assassination" case. In this case, you believe the true proposition that a prominent politician has been assassinated, but your environment is full of misleading defeating evidence, which by a fluke you never encounter. (For example, perhaps a powerful government agency has engaged in a concerted campaign of deception, planting misleading news reports in newspapers and broadcasts all over the world, falsely claiming that the politician survived the assassination attempt.) Things would only have had to be slightly different from how they actually are for you to encounter some of this defeating evidence, in addition to all the evidence that you actually have. If you had encountered this defeating evidence, your body of evidence (i) would be one that you might easily have had, in conditions that are similarly normal to the actual conditions, and (ii) would be only slightly different from the evidence that you actually have—it would have included all your actual evidence, together with this extra defeating evidence as well. If you had responded to this body of evidence, in a similarly rational way to how you responded to your actual evidence, by having some doxastic attitude

---

Footnote 14 continued

high practical credence are indeed extremely similar states. Unfortunately, however, I do not have space to repeat these arguments here. For present purposes, I hope that the reader will simply grant the assumption that outright belief and some sort of high credence are similar in the relevant way.

towards the proposition that the politician had been assassinated, you would either have had a much lower level of credence than you actually had, or else suspended judgment about the proposition altogether. So, in this case, your belief does not "adhere" to the truth.

In this case, it is not true that you might easily have a false belief about the assassination. If you encountered the misleading defeating evidence, you would suspend judgment, or have a middling level of credence like 0.5 in the proposition that the politician has been assassinated. The problem with this case is not that you might easily have believed something false. The problem is that you might too easily have *lacked* this belief because of encountering such defeating evidence.

At the same time, this version of adherence does not incorrectly rule out Nozick's (1981, p. 193) "Jesse James" case. In this case, you recognize Jesse James as he is riding past, but it is a fluke that you are looking in the right direction at the very moment when he is riding past and his mask slips. In this case, it could easily have happened that you were not looking in the right direction, or that his mask did not slip as you were looking in his direction. But if either of those things had happened, your body of evidence would have been significantly different from the evidence that you actually have: in that case, your evidence would not have properly included everything that is in your actual evidence—it would have *lacked* the distinctive kind of visual experience that triggers your ability to recognize Jesse James's face. So, the case in which you totally lack this visual experience is not sufficiently similar. Your belief that Jesse James is riding past can still count as adhering to the truth.

It is also important that this version of adherence does not require that in all of these sufficiently similar close cases, you have the *same degree* of belief that you have in the actual case. In some of these sufficiently similar close cases, you have a body of evidence that supports the target proposition *slightly less strongly* than your actual evidence. It does not prevent your actual belief from adhering to the truth if in these close cases, you have a slightly lower degree of belief in the proposition in question. Adherence fails only if in some of these close cases, you have a much lower degree of belief in, or totally suspend judgment about, the relevant proposition.[15]

Adherence guarantees that knowledge involves a *robust* connection to the truth—a connection that cannot easily be undermined by misleadingly defeating evidence. This robustness is what is invoked in *Meno* (98a), where Plato suggests that the difference between knowledge and mere true belief consists in the fact that a mere true belief is like a slave who is liable to "run away from the soul", whereas knowledge has somehow been more securely "tied down". In effect, Plato suggests, a mere true belief might too easily be *lost*, whereas knowledge is not so easily lost in this way. This suggestion does not imply that if the true belief is lost, it will be replaced by a false belief—the true belief might be replaced by doubt or uncertainty, rather than by any outright belief on the relevant topic at all. If we assume that the only relevant way of *losing* beliefs that Plato is thinking of here is these beliefs' *being defeated* or *rationally undermined*, Plato's idea that mere true belief is more easily lost than knowledge is equivalent to the kind of adherence that I am discussing here.

---

[15] This is how this version of adherence escapes the objections of Bird (2003).

Adherence seems to play a crucial role in underwriting the explanatory role of knowledge. The idea that knowledge plays a distinctive explanatory role has been stressed by Timothy Williamson. For example, in Williamson's (2000, p. 62) "burglar" case, a burglar ransacks a house all night, risking detection by staying so long. Williamson argues that there is a higher chance of the burglar's ransacking the house all night on the condition that the burglar *knows* that the house contains a diamond than merely on the condition that the burglar *truly believes* that the house contains a diamond. If the burglar merely truly believed that the house contains a diamond, then the house could have contained misleading defeating evidence, which would have come to light during the burglar's search of the house—in which case the burglar would have given up the belief that the house contains the diamond, and fled the house to escape detection. But if the burglar *knew* that the house contains the diamond, the house could not contain a mass of misleading defeating evidence in this way.

In fact, it is adherence, and not safety, that ensures that knowledge plays this explanatory role. An extremely fragile belief, which the believer would give up at the drop of a hat, can easily be extremely safe—in the sense that there is next to no chance of the believer's having a false belief on the topic in question. But if the belief is so fragile, there is also an extremely high chance of the believer's simply giving up his beliefs on this topic. In this case, the belief does not have the kind of robust connection or adherence to the truth that gives knowledge this distinctive explanatory role.

5 Why this account is (*doubly*) contextualist

As I shall explain in this section, the account that I have proposed so far is imprecise—and the most plausible way of tightening up this imprecision is by embracing a kind of *contextualism* about terms like 'know'.

According to my account, your current belief in p counts as a case of knowledge if and only if it is a case of believing correctly precisely *because* it is a case of believing *rationally*. One way in which this account is imprecise arises from the fact that rationality comes in *degrees*. Some beliefs are more rational than others. So, how rational does your belief in p have to be if you are to count as *knowing p*?

It is highly plausible that the extension of the phrase 'believing rationally' is context-sensitive. The degree of rationality that is required for it to be true to call a belief a case of "believing rationally" varies with context. In some contexts, the term is used *strictly*—and so only cases where the agent manifests a *high* degree of rationality can count as cases of the agent's "believing rationally". In other contexts, the term is used in a more *relaxed* way—and so even cases where the agent manifests a *modest* degree of rationality can count as cases of "believing rationally".

I propose that this context-sensitivity of the phrase 'believing rationally' is mirrored by a parallel context-sensitivity in the term 'know'. In some contexts, the term 'know' is used strictly, so that it is only beliefs that manifest a high degree of rationality that count as "knowledge". But in other contexts, the term 'know' is used in a more relaxed way, so that even beliefs that only manifest a modest degree of rationality can count as "knowledge".

This is in effect the kind of contextualism defended by Cohen (1999). According to Cohen, to know a proposition p, one must be justified in believing p—but justification

comes in degrees. In some contexts, the term 'know' is used strictly, so that one needs to have a lot of justification for p for it to be true in those contexts that one "knows" p, whereas in other contexts, the term is used more loosely, so that it can be true in those contexts that one "knows" p even if one only has much lower level of justification. (Like the proposal about knowledge that I am exploring here, Cohen's version of contextualism is compatible with an *internalist* conception of justification—that is, with the view that the degree to which one is justified in believing the relevant proposition depends purely on what is transpiring inside one's mind at the relevant time.)

There is also, however, a second source of imprecision in the account of knowledge that I am exploring here—a source that lies within its use of the explanatory term 'because'. I have cashed out the explanatory connection that is indicated by this 'because' in terms of what holds in all cases that are "sufficiently close" or "sufficiently similar" to the case in question. But the relevant sort of "closeness" or "similarity" comes in degrees. Some cases are closer to the case in question than others. So, how close—or how similar—does a case have to be if it is to count as sufficiently close or sufficiently similar?

Again, it seems plausible to give a contextualist answer to these questions. Sometimes, when we quantify over all "sufficiently similar" or "sufficiently close" cases, our quantifier ranges over a *large* domain of such cases: when we quantify over this large domain, it will happen relatively rarely that all the relevant cases of believing rationally (in the relevant way) are also cases of believing correctly. In other contexts, however, we might be quantifying over a much *smaller* domain of cases: when we quantify over a smaller domain, it will happen more frequently that all the relevant cases of believing rationally are also cases of believing correctly.

Larger and smaller domains of "sufficiently close" or "sufficiently similar" cases correspond to stronger and weaker interpretations of the 'because' that features in my proposed account of knowledge. So, it seems that there will be correspondingly stricter and looser ways of using the term 'know', corresponding to these stronger and weaker interpretations of 'because'.

In this way, my account *also* exhibits the kind of contextualism that has been defended by DeRose (1995) and Lewis (1996). According to both DeRose and Lewis, the extension of the term 'know' varies from one context to another, depending on which domain of possible worlds is being quantified over. The larger this domain of worlds, the fewer the number of propositions that one can truly be said to "know"; the smaller this domain of worlds, the greater the number of propositions that one can truly be said to "know".

It seems plausible to me that the term 'know' is also context-sensitive in this second dimension as well: every knowledge-attribution implicitly quantifies over the domain of sufficiently close or sufficiently similar possible cases, and this domain of possible cases varies with the context. This is not to say that I am committed to everything that contextualists like Cohen, DeRose, and Lewis have defended.[16] In particular, I do not

---

[16] DeRose (1995, p. 34) says that for one to know p, one's "belief as to whether p is true" must "match the fact of the matter as to whether p is true, not only in the actual world, but also at the worlds sufficiently close to the actual world." While this formulation is not exactly wrong, it is misleading. This formulation suggests

intend to endorse their position that the context-sensitivity of 'know' plays a significant role in defusing the arguments that seem to support radical scepticism. But it seems to me that the best way to resolve the two forms of imprecision that I noted in my account of knowledge is by accepting the contextualists' central thesis—that the extension of the term 'know' varies with the context in which it is used, sometimes picking out a stronger notion (so that in these contexts, there are relatively few propositions that we can truly be said to "know") and sometimes a weaker notion (so that in these other contexts, there are many more propositions that we can truly be said to "know").[17]

6 Are there counterexamples to this account?

Our grasp of the concept of knowledge generates a rich body of intuitions about cases. In this section, I shall investigate whether any of these intuitions provide counterexamples to my proposed account of knowledge.

I shall start with the question of whether there are counterexamples to the *safety* condition on knowledge. Many objections to safety target the claim that safety is *sufficient* for knowledge. But that claim is not being defended here: according to the account of knowledge proposed above, safety is a necessary but not sufficient condition for knowledge.

However, some philosophers deny that safety is even *necessary* for knowledge. For example, Sosa (2007, pp. 29, 41) has argued that safety is not necessary for knowledge. It turns out, however, that these objections to safety focus on different and at least slightly stronger conceptions of safety than the one that I have advocated here.

According to the kind of safety that I have defended above, the relevant cases—in which the thinker must not have a false belief—are cases that are similar (a) with respect to what makes the actual case a case of *rational* thinking, and (b) with respect to the way in which, and the degree to which, the actual case counts as *normal*. (In discussing the "adherence" condition in Sect. 4, I also made some more suggestions about these cases: they must be cases in which (i) one has a similar body of evidence, which one could easily have had in conditions that were similarly normal to the actual case, and (ii) one responds to this evidence by having some doxastic attitude towards the case's target proposition.) When none of these close cases involve a false belief, the actual case is safe. This kind of safety can be called "rationality-and-normality-relative safety", or "RN-safety" for short.

Footnote 16 continued

that this condition could be met because the relevant domain of close worlds includes some worlds at which p is not true, and at all such worlds, one *believes* that p is not true. But in fact, knowing p does not require believing that p is *not* true at close worlds where p is not true. (There are plenty of propositions—such as existential propositions like 'There is a spider in this room'—that are much easier to know than their negations!) According to my proposal, what is required is that, at every case in the relevant domain of cases, (a) the case's target proposition is *true*, and (b) one either believes or at least has high credence in that proposition. So, in fact, the relevant domain must not contain *any* case where the case's target proposition is not true.

[17] For a more extensive discussion of contextualism, and a rebuttal of some of the objections that have been raised against it by such philosophers as Stanley (2005), see Wedgwood (2008).

The kinds of safety that Sosa criticizes as unnecessary for knowledge are in fact stronger than RN-safety. These stronger kinds of safety are what Sosa calls "outright safety" and "basis-relative safety". For your belief in a proposition p to be "outright safe", it must not be the case in any sense that you "might too easily" have had a false belief in p; for your belief to be "basis-relative safe", it must have "some basis that it would not easily have had unless true, some basis that it would (likely) have had only if true" (29).

There are cases where your belief is RN-safe, but not "outright safe" or "basis-relative safe". Suppose that you could easily have held a belief in a false proposition on the same kind of basis on which you actually believe p; but that if that had happened, it would have been because either (a) your rational dispositions were impaired in a way in which they actually were not impaired, or else (b) your external conditions were abnormal in a way in which they were actually not abnormal. For example, suppose that you form a perceptual belief in a perfectly ordinary way; but as it happened, the evil demon was all prepared to start deceiving you today (in which case the content of the perceptual beliefs that you would have formed would have been false), and it was only through an extraordinary series of freak accidents that the demon's plans failed. Then it could easily have happened that you were in the highly abnormal circumstances of being deceived by the demon—even though in fact, because the demon's plans fell through, your actual circumstances are perfectly normal. In this case, as Sosa points out, your perceptual belief does not exhibit basis-relative safety (nor *a fortiori* outright safety), but it surely could still be a case of knowledge.

However, cases of this sort are not counterexamples to the claim that knowledge requires RN-safety, since in all these cases, the belief in question could still be RN-safe. To judge whether a belief is RN-safe, cases where the believer is thinking in a less rational way than in the actual case are irrelevant, as are cases in which the believer's circumstances are less normal than they actually are. For RN-safety, what is necessary is that the believer should not have a false belief in any case in which the thinker is thinking in a similarly rational way, in similarly normal external circumstances. The cases that Sosa describes are not counterexamples to my claim that knowledge requires RN-safety.

As I explained in Sect. 4 above, my account of knowledge implies that knowledge requires adherence as well as safety. Several philosophers have objected to adherence, by describing cases in which we intuitively have knowledge, but in which there are allegedly close sufficiently similar cases in which the case's target proposition is true but not believed. For example, Sosa (2002, p. 274) brings up the following case:

> One can know that one faces a bird when one sees a large pelican on the lawn in plain daylight even if there might easily have been a solitary bird before one unseen, a small robin perched in the shade, in which case it is false that one would have believed that one faced a bird.

In a similar vein, Kripke (2011, p. 178) produces the following counterexample:

> Suppose that Mary is a physicist who places a detector plate so that it detects any photon that happens to go to the right. If the photon goes to the left, she will have no idea whether a photon has been emitted or not. Suppose a photon is emitted,

that it does hit the detector plate (which is at the right), and that Mary concludes that a photon has been emitted. Intuitively, it seems clear that her conclusion indeed does constitute knowledge. But is Nozick's fourth condition satisfied? No, for it is not true, according to Nozick's conception of such counterfactuals, that if a photon had been emitted, Mary would have believed that a photon was emitted. The photon might well have gone to the left, in which case Mary would have had no beliefs about the matter.

However, it seems to me that both of these cases are like the "Jesse James" case. The close possible cases in which one lacks the belief are not sufficiently similar with respect to what makes one's actual belief rational, because the evidence that the agent has in those cases is significantly different from the evidence of the actual case. What is incompatible with adherence is the existence of close cases in which the believer has a very similar body of evidence—such as a body of evidence that includes all of the believer's actual evidence but also some additional defeating evidence as well—and the proposition is true, but the believer rationally has a much lower degree of belief in the proposition in question. The cases described by Sosa and Kripke do not involve sufficiently similar evidence of this kind.

Setiya (2013, p. 91f.) has suggested a more general objection to adherence: "I can know the truth by a method whose threshold for delivering a verdict is extremely high, so high that it virtually always leaves me agnostic." In fact, however, there seem to be two very different kinds of case that satisfy Setiya's description.

First, suppose that one believes a mathematical truth p by the following strange method: when one produces a mathematical proof of a proposition, one tosses a coin four times, and then believes the proposition that one has proved only if the coin lands heads all four times—otherwise, one suspends judgment about the proposition. Here it could easily happen that the coin lands tails on one of the tosses, in which case one would have been agnostic about the proposition. In this case, however, I am not convinced that one really does know the proposition in question. The explanation of one's belief seems to make it clear that one's belief is wildly irrational: one holds the belief only because of the utterly irrelevant evidence about the outcome of the coin toss. If I am right that knowledge requires rationality, then it would be correct to deny that this irrational belief is a case of knowledge.

Secondly, one's method might be a perfectly rational method, but it might have a high threshold for delivering a verdict because it requires having a rare and unusual kind of evidence. But then the case is just like the cases of Sosa and Kripke that we have just discussed: it could indeed easily happen that one lacks the unusual kind of evidence that supports the proposition, but the fact that in such cases one would lack the belief does not show that it does not adhere to the truth. The belief adheres to the truth because in cases in which one has a very similar body of evidence, which one might easily have had (and the proposition in question is true), one has a belief that is at most only *slightly* less confident than one's actual belief. Setiya's general objection

does not point to any convincing case of knowledge in which one's belief does not adhere to the truth in this sense.

7 The significance of the concept of knowledge

Clearly, this version of the virtue theory of knowledge can avail itself of something very similar to Sosa's (2015, p. 40) answer to the question that Plato raises in *Meno*: Why does knowledge strikes us as more valuable than mere true belief? This account agrees with Sosa's that knowledge is in a sense an achievement, or a manifestation of virtue, while mere true belief is in a way sheer good luck.

However, interpreting knowledge as an admirable achievement does not explain another striking fact—the fact that verbs like 'know' and its equivalents in other languages are among the commonest verbs in daily use. As I pointed out in earlier work (2002, p. 289):

> we almost *never* aim to have true belief without at the same time aiming to know. Indeed, as Bernard Williams pointed out (1978, p. 37), we would not ordinarily suppose that 'aiming to get to the truth' and 'aiming to know' describe different aims at all. On the contrary, we would ordinarily assume that they describe *the very same aim*.

In my earlier work, I endorsed the explanation that Bernard Williams (1978, pp. 38–45) gave of the relation between "aiming at the truth" and "aiming to know". As I put it:

> The reason why we would ordinarily take these two phrases as describing the very same aim is this: a rational thinker cannot pursue the aim of believing the truth and nothing but the truth, except in a way that, if it succeeds, will result in knowledge.

According to my account, knowledge is effectively identical to *success* in the *rational pursuit* of the *truth*. Any rational agent who aims to get to the truth must rationally pursue that aim, and must also aim to succeed in that pursuit. (This is not a special feature of aiming at the truth: any rational agent who aims at anything must rationally pursue the aim, and aim to succeed in that pursuit.) So, it may seem that my account agrees with Williams's about the relation between "aiming at the truth" and "aiming to know".

On reflection, however, there seems to be a gap in this explanation. My account of knowledge is offered as an account of *what it is* for a thinker to know a proposition— that is, as an account of the *state* of knowledge. It is not offered as an account of the *concept* of knowledge—that is, as an account of how ordinary thinkers *represent* or *conceive of* that state in ascribing knowledge to various subjects. To explain why we would ordinarily think that "aiming to get to the truth" and "aiming to know the truth" are one and the same aim, we need an account of how we ordinarily *think* of knowledge—that is, an account of the concept of knowledge—not just an account of the state of knowledge.

Might it be correct simply to say that possession of the concept of "knowledge" requires an "implicit grasp" of the account of knowledge that I have given above? There are reasons to doubt whether it would be correct to say this. While the term 'know' is an extraordinarily common term in English, terms like 'rational' and 'justified' are significantly less common. So, it is far from obvious that ordinary thinkers on the street have an implicit grasp of the account that I have given. Indeed, it looks as if it might be quite possible for a simple thinker to master the concept of "knowledge" even if they do not possess the concept of "rationality" at all.

So, what is it to possess the concept of "knowledge"? I shall assume that the following conditions are all essentially involved in our possession of the concept:

a. Knowledge is conceptually *factive*: that is, possessing the concept of "knowledge" constitutively requires an ability to infer from 'S knows p' to p.
b. Knowledge conceptually *entails belief*: that is, possessing the concept of "knowledge" constitutively requires an ability to infer from 'S knows p' to 'S believes p'.
c. When we judge that a thinker S knows p, we are in a way *endorsing S's* belief about p: as Craig (1990, p. 11) put it, we are "flagging" S as "an approved source of information" about p.
d. It is a conceptual truth that knowledge is in principle *explicable*: if S knows p, there is in principle an answer to the question "How does S know p?" An answer to this question gives at least part of an explanation of why S is right about p—that is, of why it is that S has a correct belief about p.[18]

At the same time, if my account of the state of knowledge is correct, a good account of the concept of "knowledge" must explain why the concept picks out or refers to the state that is analysed by my account of the state of knowledge.

To solve this problem, I propose that there is a kind of *endorsement* that we can have towards a belief—a kind of endorsement that is appropriate if and only if the belief is rationally held. A thinker can have this kind of endorsement towards a belief even if the thinker does not explicitly judge that the belief is rational—and even if the thinker does not even possess the concept of "rationality". Nonetheless, this kind of endorsement has an essential relationship to rationality, because it is appropriate to have this kind of endorsement towards a belief if and only if the belief is rational.

Then we could say: To think of someone as knowing p is to endorse the believer's belief in p in this way, and to take one's endorsement as pointing to an explanation of why the believer is right about p. What exactly is it for one to "take" one's endorsement of a belief in p to "point to" an explanation of why the believer is right about p? I tentatively suggest that it is for one implicitly to believe of one's endorsement that

---

[18] An anonymous reviewer for this journal has asked whether it might not be coherent to suppose that an omniscient God knows all truths, even though there is no answer to the question, "How does God know?" But many theological accounts of omniscience have attempted to answer this question. (For example, perhaps God knows because he is infallibly aware of his own will and of his omnipotence; or perhaps God's omnipresence makes him omni-percipient; or perhaps our limited minds cannot grasp the truth about how God knows….) At all events, a theologian who explicitly claims that God knows everything even though there is absolutely no answer at all to the question "How does God know?" does seem to me to verge on incoherence.

there are facts that make it appropriate, and those facts feature in an explanation of why the believer is right about *p*.

Much more investigation is required if we are adequately to clarify and defend this account of the concept of "knowledge". However, it seems that something like this account could fill the gap in the account that I proposed earlier of why "aiming to get to the truth" and "aiming to know the truth" strike us as the very same aim—so long as we can supplement our account with an explanation of why no rational agent will have an aim without in the relevant way "endorsing" her own pursuit of the aim. Clearly, it also cannot be rational to pursue an aim, while aiming for one's pursuit to realize the aim through a lucky accident. If one rationally pursues an aim, one must also aim for one's pursuit to succeed at achieving the aim. In this way, the account of knowledge that I have proposed here looks as if it can be harmonized with a plausible account of the significance of the concept of knowledge.

In conclusion, I would like to note that the account of knowledge that I have proposed here tries to build on a large number of insights of other epistemologists who have worked on the nature of knowledge over the years: virtue theories like Sosa's "aptness" theory; "safety" theories like those of Williamson and the earlier Sosa; the idea of a belief's "adhering" to the truth, which was defended by Nozick; and the ideas of the contextualists like Cohen, Lewis, and DeRose; and many other insights of other epistemologists as well. All of these theorists were right about an important part of the picture. Indeed, even JTB was not radically wrong as an account of knowledge. According to my proposal, in effect, every case of knowledge is a case of TB-*because*-JB. In this way, all that my proposal does is simply to replace JTB's bare conjunction of TB and JB with an explanatory connection. All of these epistemologists have made progress towards understanding the nature of knowledge. To complete the task, all that is required is to connect their insights together.[19]

References

Bird, A. (2003). Nozick's fourth condition. *Facta Philosophica*, *5*, 141–151.

Bonjour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy*, *5*, 53–73.

Cohen, S. (1999). Contextualism, skepticism, and the structure of reasons. *Philosophical Perspectives*, *13*, 57–88.

Conee, E., & Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies*, *89*(1), 1–29.

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

Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.

DeRose, K. (1995). Solving the skeptical problem. *Philosophical Review*, *104*(1), 1–52.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Goldman, A. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, *73*(11), 771–791.

Harman, G. (1968). Knowledge, inference, and explanation. *American Philosophical Quarterly*, *5*(3), 164–173.

Kripke, S. (2011). *Philosophical troubles*. Oxford: Clarendon Press.

Lackey, J. (2007). Why we don't deserve credit for everything we know. *Synthese*, *158*(3), 345–361.

Lackey, J. (2008). Knowledge and credit. *Philosophical Studies*, *142*(1), 27–42.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, *74*(4), 549–567.

Nozick, R. (1981). *Philosophical explanations*. Cambridge: Harvard University Press.

Pettit, P. (1993). *The common mind: An essay on psychology, society, and politics*. Oxford: Oxford University Press.

Setiya, K. (2013). *Knowing right from wrong*. Oxford: Clarendon Press.

Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, *13*, 141–153.

Sosa, E. (2002). Tracking, competence, and knowledge. In P. K. Moser (Ed.), *The Oxford handbook of epistemology*. Oxford: Oxford University Press.

Sosa, E. (2007). *A virtue epistemology*. New York: Oxford University Press.

Sosa, E. (2015). *Judgment and agency*. Oxford: Oxford University Press.

Stanley, J. (2005). *Knowledge and practical interests*. Oxford: Oxford University Press.

Turri, J. (2010). On the relationship between propositional and doxastic justification. *Philosophy and Phenomenological Research, 80*(2), 312–326.

Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, *16*, 267–297.

Wedgwood, R. (2006). The normative force of reasoning. *Noûs*, *40*(4), 660–686.

Wedgwood, R. (2008). Contextualism about justified belief. *Philosophers' Imprint, 8*(9), 1–20. www.philosophersimprint.org/008009/.

Wedgwood, R. (2012a). Outright belief. *Dialectica, 66*(3), 309–329.

Wedgwood, R. (2012b). Justified inference. *Synthese*, *189*(2), 273–295.

Wedgwood, R. (2013). The right thing to believe. In T. Chan (Ed.), *The aim of belief* (pp. 123–139). Oxford: Oxford University Press.

Wedgwood, R. (2017). *The value of rationality*. Oxford: Oxford University Press.

Williams, B. (1978). *Descartes: The project of pure enquiry*. Harmondsworth: Penguin.

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