

A VIRTUE RELIABILIST ERROR- THEORY OF DEFEAT

Jaakko Hirvelä

forthcoming in Erkenntnis

ABSTRACT: Knowledge defeat occurs when a subject knows that p , gains a defeater for her belief, and thereby loses her knowledge without necessarily losing her belief. It's far from obvious that externalists can accommodate putative cases of knowledge defeat since a belief that satisfies the externalist conditions for knowledge can satisfy those conditions even if the subject later gains a defeater for her belief. I'll argue that virtue reliabilists can accommodate defeat intuitions via a new kind of error theory. I argue that in cases where the subject holds dogmatically onto her belief in the face of an apparent defeater, her belief never qualified as knowledge, since the belief was not gained via an exercise of her epistemic virtues. In cases where the subject suspends her judgment upon receiving the putative defeater her original belief might have qualified as knowledge, but crucially, in such cases knowledge is lost due to loss of belief, rather than due to the epistemic force of the defeater. Therefore, knowledge defeat isn't a genuine phenomenon even though there are no cases where a subject knows what she originally believed after receiving the putative defeater.

Keywords: Defeat; virtue reliabilism, cognitive integration, knowledge, justification

1. INTRODUCTION

Knowledge defeat is said to occur when a subject knows that p , then gains a putative defeater for her belief, and thereby loses her knowledge that p without necessarily losing her belief that

p or any relevant evidence. Accommodating the phenomenon of knowledge defeat isn't easy for externalist theories of knowledge.¹ Indeed, some of the best arguments that have been mounted against externalism rely on putative cases of knowledge defeat.² Here are two representative cases:

Red light

At t_1 Yen comes to know that the wall in front of her is red via perception in optimal conditions. At t_2 Yen's trusted friend Triss tells her that the wall is illuminated by red light that would have made the wall look red even if it had been of some other colour.³

Feint

At t_1 Ciri remembers that she has a fencing session at four o'clock. There is nothing wrong with her memory, and her belief qualifies as knowledge. At t_2 Ciri's friend Zoltan tells her that he just came from the fencing hall and saw on the calendar that Ciri's fencing lesson is at five o'clock.⁴

The defeatist intuition is that Yen and Ciri *must* lose their knowledge after acquiring the defeater, even if they hold onto their respective beliefs. The defeatists claim that acquiring a defeater for a belief that qualifies as knowledge *suffices in itself* to defeat the knowledge-status

¹ Following Lasonen-Aarnio (2010, p. 18), by externalist theories of knowing I mean theories that don't require satisfying an internalist justification condition that could be used to explain knowledge defeat. Goldman (1979), Nozick (1981), Williamson (2000), Lasonen-Aarnio (2010), and Pritchard (2012), for example, sport such theories of knowing (to the extent that they can be said to hold 'theories' of knowledge in the first place).

² See for instance Bonjour (1980, pp. 59-60). Beddor (2015) has argued that process reliabilists cannot account for defeat.

³ Lasonen-Aarnio (2010) discusses a similar case.

⁴ This case features a rebutting defeater, which aims to undermine the truth of the subject's belief. **Red light**, in contrast, features an undermining defeater, that aims to undermine the positive epistemic support that the subject has for her belief, without necessarily undermining the truth of the subject's belief. See Pollock (1987, p. 485) for the distinction of undermining and rebutting defeaters.

of the belief. Externalist theories of knowledge have a difficult time accommodating the defeatist intuition because it's far from obvious that Yen and Ciri cannot continue to hold onto their beliefs in a reliable way, given that the beliefs were reliably formed and maintained before they were made aware of the defeaters. For example, if knowledge is safe belief, that is, a belief that could not easily be false given the basis on which it's held, then assuming that Yen's and Ciri's beliefs were safe at t_1 it's hard to see why they could not be safe at t_2 .⁵ After all, simply acquiring new information need not change the modal profile of the belief.⁶ Yen and Ciri can simply fail to rebase their beliefs on their new evidence. If Yen is asked why she believes that the wall is red, she will reply that she *sees* that the wall is red. She won't reply that she knows that the wall is red because she sees that the wall is red even though the wall is illuminated by red light that would have made it look red even if it was not red. Of course, if Yen and Ciri do rebase their beliefs on their new evidence, then their beliefs might become unsafe, since they might have formed a different belief, which would have been false. But whether Yen and Ciri rebase their beliefs is a contingent matter, and therefore their knowledge need not be defeated by the putative defeater, contra the defeatist intuition (Lasonen-Aarnio, 2010).⁷

Therefore, it isn't easy to see how externalist theories of knowledge could accommodate putative cases of knowledge defeat. An option that used to be popular was to add a no-defeaters

⁵ The safety condition has been put forth by Sainsbury (1997), Sosa (1999), Williamson (2000), Pritchard (2005), and Hirvelä (2019a), among others. For critique of the safety condition, see Comesana (2005), Neta and Rohrbaugh (2004), Miracchi (2015), and Paterson (forthcoming).

⁶ Lasonen-Aarnio (2010) has argued that beliefs that are retained in defeat cases need not be unsafe, and hence can amount to knowledge. I have (2017) argued that the fact that an epistemic peer disagrees with you doesn't affect the safety of your belief. If knowledge is safe belief, then peer disagreement cannot undermine knowledge. Goodman and Salow (2018) have modeled knowledge in a broadly safety-theoretic framework within which misleading evidence can defeat knowledge. It's worth to note that their model also validates the KK-principle, which standard formulations of safety don't validate. Hence the safety condition Goodman and Salow endorse differs significantly from the mainstream version. I thank an anonymous reviewer at *Erkenntnis* for alerting me to Goodman and Salow's work.

⁷ Note also that if we were to think that what one knows is always part of one's evidence, then if Yen and Ciri don't lose any evidence upon receiving the misleading defeater, their evidence at t_2 will still conclusively support their original beliefs because knowledge is factive.

clause to the externalist theory of knowledge.⁸ Those who have been unwilling to add a seemingly *ad hoc* no-defeaters clause to their accounts of justification or knowledge have aimed to accommodate intuitions of knowledge defeat via error-theories.

Lasonen-Aarnio (2010, 2014, forthcoming-b) has argued that knowledge can sometimes be retained in putative cases of knowledge defeat. Our negative assessment of subjects who retain their knowledge in such cases is explained by the fact that they are manifesting bad dispositions, dispositions that would in general be manifested in cases of ignorance, rather than in cases of knowledge. Baker-Hytech and Benton (2015, p. 57) have argued that if knowledge is the norm of belief, then the apparent irrationality of subjects who retain their beliefs in face of misleading evidence can be explained by the fact that such subjects violate a guidance norm that is generated by the knowledge norm of belief.⁹

While I am very sympathetic to both Lasonen-Aarnio's proposal and to Baker-Hytech and Benton's account, I wish to sketch a new kind of error-theory that falls directly out of virtue reliabilism. The error-theory I propose differs significantly from the earlier ones. According to it, in cases where the subject dogmatically clings onto her belief she never knew to begin with, or did not acquire a putative defeater in the first place, while in cases where the subject suspends judgment she might have known, but doesn't any more since she lacks the relevant belief. It's an error theory for two reasons. Firstly, in some cases of putative knowledge defeat we mistakenly think that the subject had knowledge to begin with. Secondly, according to the view putative defeaters cannot on their own defeat knowledge. The defeatist intuition is explained by the fact that there are no cases where a subject knows that p at t_1 , and retains her knowledge of p after having received a putative defeater for p at t_2 .

⁸ Goldman has aimed to deal with putative defeat cases by adding a no-defeaters clause to his theory of justification, and hence to his theory of knowledge. For different ways in which a no-defeaters clause can be added to process reliabilist theories see Goldman (1979; 1986, pp. 111-112) and Lyons (2009, 2016). For critique of these proposals, see Beddor (2015).

⁹ See Brown (2018, p. ch. 5) for discussion of these strategies to explain away knowledge defeat.

Some readers might think that the error-theory provided is too radical. These readers are invited to see this paper as offering an argument *against* virtue reliabilism, since the error-theory I present falls directly out of the main tenets of virtue reliabilism. My sole aim here is to examine what consequences virtue reliabilism has for defeat.

In what follows, I'll focus on cases like **Red light** and **Feint** that involve so-called doxastic or mental-state defeaters. All doxastic defeaters are beliefs. For instance, in 'Red light' the defeater that Yen has is her belief that [the wall is illuminated by red light that would have made the wall look red no matter its actual colour]. I'll set aside cases that feature propositional or normative defeaters (Lackey, 1999). A propositional defeater for the belief that p is a true proposition such that if S were to believe it, then S wouldn't know that p . A normative defeater is a propositional defeater that the subject should have believed. There are two reasons why I limit the scope of inquiry to the potential epistemic force of doxastic defeaters. Firstly, cases featuring doxastic defeaters are the most plausible cases of knowledge defeat. If it turned out that doxastic defeaters are void of epistemic power, as I hope to show, then there is reason to think that propositional and normative defeaters are void of epistemic power too. Secondly, I think that cases of propositional and normative defeat are highly contentious. In my mind it's better not to use such cases when evaluating a theory. Henceforth all talk of defeaters refers to doxastic defeaters.

This essay is structured as follows. In the next section I lay out some key ideas of virtue reliabilism. In the third section I spell out under what conditions a belief can function as a defeater. In the fourth section I examine whether knowledge defeat is a genuine phenomenon, under the assumption that knowledge is always the product of one's cognitive abilities. In the fifth section I briefly compare my account to other virtue theoretic solutions.

2. VIRTUE AND COHERENCE OF CHARACTER

The central thesis of virtue reliabilism is that knowledge requires that one's cognitive success must be attributable to one's cognitive character. Some virtue epistemologists see this central

thesis as giving both necessary and sufficient conditions for knowledge.¹⁰ Others think that it provides only a necessary condition for knowledge.¹¹ The central thesis can be interpreted in various ways. A cognitive success can be understood either as the acquisition or maintaining of a *true belief*, or as the acquisition or maintaining of *knowledge*. The former views belong to the classical tradition of analyzing knowledge in terms of true belief plus some other conditions.¹² The latter views belong to the knowledge first movement, championed by Williamson (2000).¹³ Another aspect in which the central thesis is ambiguous is on the question when a cognitive success is attributable to the subject's cognitive abilities. According to Greco (2010), the truth of a subject's belief is attributable to her cognitive character just in case the fact that she believes out of cognitive character is part of the most salient causal explanation why she acquired a true, rather than a false belief. In other words, one's cognitive character has to be an important part of the best causal explanation for one's cognitive success in order for one to know. According to Sosa (2007, 2009), one's cognitive success is attributable to one's cognitive character just in case one's cognitive success is a manifestation of the cognitive abilities that make up one's cognitive character. Thus Sosa's seeks to understand the attribution relation in terms of a more general metaphysical relation, namely, as the manifestation of a disposition. Many have preferred Sosa's account to Greco's, probably because Sosa is able to side-step some counterexamples that Greco's account seems susceptible to.¹⁴

¹⁰ For some of the most prominent statements of the view, see Sosa (2007, 2009, 2011, 2015), Greco (2010), and Zagzebski (1996).

¹¹ See, Pritchard (2012) and Kelp (2013).

¹² Sosa, Greco, and Zagzebski belong to this tradition.

¹³ Miracchi (2015) Silva (2017) and Kelp (2017) have argued for knowledge first virtue epistemology. See Hirvelä (2019b) for critique of knowledge first virtue epistemology.

¹⁴ For instance, Turri (2011), Littlejohn (2014), and Kelp (2017) invoke the notion of manifestation of a disposition in their understanding of the attribution clause. Lackey (2007, 2009) argues on the basis of testimonial cases of knowledge that knowing doesn't require that one's cognitive success be attributable to one's cognitive character. It would seem that Sosa (2007, pp. 95-96; 2011, p. 87) has the means to deal with Lackey's objections, but it's not clear whether Greco does. In fact Greco (2012) has changed his view in light of Lackey's apt criticism. I think that Greco's new proposal is better suited to deal with

Here we need not be concerned with these issues. The argument I offer doesn't depend on how one understands cognitive success, nor on how we flesh out the attribution clause. In fact, virtue-theoretic views that don't invoke the attribution relation, but merely require that one's belief has to be the product of one's cognitive abilities, also fall under the scope of the views I wish to discuss.¹⁵ What all of these views share is the idea that knowledge requires the use of cognitive abilities. In order to make use of a cognitive ability one must possess that ability. But under what kind of conditions does one possess a certain ability?

In a broadly Aristotelian spirit, virtue epistemologists think that a reliable doxastic disposition can count as a cognitive virtue only if it's a proper part of one's virtuous *epistemic character* (Greco, 1999, p. 287; 2010, p. 150; Palermos, 2014, p. 1940; Pritchard, 2012, p. 262).^{16, 17} This is what differentiates virtue reliabilism from process reliabilism. Virtue reliabilists require that the reliable processes be properly grounded in the subject in order to be knowledge-conducive. According to them, not all reliable doxastic dispositions count as cognitive abilities. If Alvin has a brain lesion that causes him to believe that he has a brain lesion, his belief is the product of an extremely reliable doxastic disposition, but it's not a product of his cognitive abilities, because the brain lesion isn't a part of Alvin's cognitive character (Breyer & Greco, 2008, p. 174; Greco, 2010, p. 151; Palermos, 2014, p. 1938).¹⁸

Lackey's objections. For discussion of different ways to understand credit, see Hirvelä and Lasonen-Aarnio (forthcoming).

¹⁵ See Hirvelä (2018, 2019a) and Beddor and Pavese (2020) for a virtue-theoretic view that doesn't invoke the attribution relation. The virtue-theoretic condition that Pritchard (2012) endorses does demand that the agent's cognitive success be of credit to her, and hence is logically stronger.

¹⁶ Knowledge first virtue reliabilists think that the relevant cognitive abilities are abilities to know, whereas those virtue reliabilists who have reductive ambitions understand such abilities as abilities to gain or maintain true beliefs. In what follows we can remain neutral on this score.

¹⁷ Is the notion of character essential in virtue epistemology? Perhaps not. Sylvan (2017), drawing on the work of Thomson (1997) and Hurka (2006), develops an intriguing virtue responsibilist view that takes act-attaching virtue properties to be fundamental, rather than character-attaching virtue properties. This kind of virtue theory is outside the scope of my argument.

¹⁸ This case is due to Plantinga (1993, p. 199).

But under what conditions is a reliable doxastic disposition a proper part of one's cognitive character? At least three conditions have been proposed by virtue reliabilists; (i) that the disposition is stable, (ii) that it's not strange, and, (iii) that it's integrated into the subject's cognitive character (Greco, 2010, p. 150). The central idea behind these conditions is to ensure that in order for a doxastic disposition to be a part of one's cognitive character it has to be the *agent's* disposition. Beliefs that are products of such cognitive abilities are in a sense *owned* by the subject, in that she is responsible for those beliefs and can be properly blamed or credited for having those beliefs. I'll focus on condition (iii), since it seems to be the most central one, and is more widely endorsed than conditions (i) and (ii).¹⁹

What suffices for cognitive integration varies from case to case. In some extreme cases, like in the brain lesion case, reflective endorsement of the truth-conduciveness of the disposition might be required (Pritchard, 2010). If, for instance, Alvin went to see a doctor who told him that he suffers from an extremely rare brain lesion that causes one to believe that one suffers from a brain lesion, the doxastic disposition generated by the brain lesion could become a part of Alvin's cognitive character. But this kind of reflective endorsement is almost never required in more mundane cases. Doxastic dispositions that are innate, or otherwise naturally developed, are integrated into our cognitive system via subconscious mechanisms in virtue of constantly confirming each other's outputs. Consider for example the following description of Edgar's afternoon: Edgar sees a beautiful pint of ale and can smell the overwhelming aroma of the hops. He can feel the cold glass in his hand and sipping the beer, finds delightful notes of pine, citrus and tropical fruits. Pricking his ears he can even hear the dense head slowly dissolving, and thinks: "I'm drinking ale today". All of these experiences confirm to Edgar that there's a pint of ale on the table.

In Edgar's case all of his sensory modalities part-take in confirming a single proposition. Of course this isn't always the case. For many sensible qualities it applies that they can be sensed

¹⁹ Breyer and Greco (2008); Greco (2010); Pritchard (2012) and Palermos (2014) are explicit in endorsing (iii).

directly only via some particular sense modality. No one can hear the redness of the wall. Its redness can only be seen. However, many of our experiences are multi-modal in that multiple sense modalities are responsible for our phenomenological state. And it's not just the case that our sense modalities confirm the outputs of each other. Rather, in many cases our sense modalities affect the outputs and operation of our other sense modalities.²⁰

A minimal externalist condition for cognitive integration is that the cognitive abilities act in concert with each other. Greco (2010, p. 152) writes that “cognitive integration is a function of cooperation and interaction, or cooperative interaction, with other aspects of the cognitive system.” Palermos (2014, pp. 1941-1942) holds that “the only necessary and sufficient condition for a process to count as knowledge-conducive is that it cooperatively interacts with the rest of the agent's cognitive character. [The] process of cognitive integration gives rise to a coherentist effect both on the level of processes (how the beliefs are generated) and on the level of content (how the beliefs themselves combine).” Pritchard (2010, pp. 147-148) holds that a doxastic disposition D is integrated to the subject's cognitive character only if beliefs gained via a D have cohered with the beliefs formed via the subject's other cognitive abilities, and that if they had not, then the subject would have responded accordingly. Sosa argues also that knowledge never arises purely from one faculty, but from the interplay of cognitive faculties.²¹ He writes:

Note that no human blessed with reason has merely animal knowledge of the sort attainable by beasts. For even when perceptual belief derives as directly as it ever does from sensory stimuli, it is still relevant that one has *not* perceived the signs of contrary testimony. A reason-endowed being automatically monitors his background information and his sensory input for contrary evidence and automatically opts for the most coherent hypothesis even when he responds most directly to sensory stimuli. [...] The beliefs of a *rational* animal hence would

²⁰ For an overview of recent empirical work on multi-modality see O'Callaghan (2012).

²¹ Thanks to Kurt Sylvan for pointing me towards relevant passages in Sosa's work.

seem never to issue from *unaided* introspection, memory, or perception. For reason is always at least a silent partner on the watch for other relevant data, a silent partner whose very *silence* is a contributing cause of the belief outcome. (Sosa, 1991, p. 240)

This kind of minimal integration doesn't require any perspective on the truth conduciveness of the dispositions. The only thing that is required is that the doxastic dispositions that make up one's virtuous cognitive character are not acting in conflict with each other. Hence we are able to lay down the following condition for minimal cognitive integration:

INTEGRATION: Subject S's doxastic disposition D is integrated with her cognitive character only if D would act in concert with the set of doxastic dispositions D* that together with D make up S's cognitive character if D were triggered while both D and D* are in appropriate conditions.

Given INTEGRATION, a reliable doxastic disposition can qualify as a cognitive ability just in case it acts, or would act, in concert with one's cognitive character, while in appropriate conditions.²² According to virtue reliabilism only beliefs gained via cognitive abilities can have

²² Note that it is not enough that it would be merely probable that the disposition acts in concert with one's cognitive character. Virtue reliabilists have at least two reasons why they should not opt for a weaker reading of 'would' in INTEGRATION. First, INTEGRATION demands that the relevant dispositions act in concert with each other when triggered while in *appropriate* conditions. Many virtue reliabilists understand appropriate conditions in terms of *normal* conditions, or conditions that are otherwise suitable for the exercise of the ability in question (Greco 2010, Sosa 2010, Beddor & Pavese 2020). Therefore, INTEGRATION is already effectively weakened in that it requires only that the dispositions would normally act in concert with each other. Second, if a doxastic disposition D could be a part of S's cognitive character even though it would be merely probable that it acts in concert with S's cognitive character while in appropriate conditions, then the performances that D would issue which were in tension with S's cognitive character would be attributable to S, since they would be manifestations of her cognitive abilities. But within the literature on attributability, even outside virtue epistemology, it is commonplace to think that an act is attributable to an agent "just in case it expresses the agent's deep self" Shoemaker (2015, p. 59). But how could a performance that is out of character express, or reveal the agent's cognitive character? I contend that it could not. I would like to thank an anonymous reviewer at *Erkenntnis* for raising this issue.

positive epistemic statuses like justification or knowledge. Knowledge and justification require a kind of coherence of one's cognitive faculties. For a subject to be eligible for such normative statuses she must keep her cognitive home in order.²³

One might object that INTEGRATION is too strong. Even though I know by testimony that the Müller-Lyer lines are equally long I still see them as of different lengths. When in the grips of the Müller-Lyer illusion my eyesight doesn't seem to act in concert with the other doxastic dispositions that make up my cognitive character. But here it's important to note that I don't form the belief that the lines are of different lengths on the basis of my perceptual experience when I know that they are of the same length. The fact that I don't form the belief is evidence that my eyesight is acting in concert with my cognitive character, since the knowledge that I've gained through my other cognitive faculties prevents me from forming a belief that corresponds to the experiential-state generated by my eyesight.

It would be good if we could say more about what it takes for two doxastic dispositions to act in concert with each other. Sadly virtue reliabilists have been largely silent on this issue. What seems clear, however, is that two doxastic dispositions can act in concert with each other just in case the dispositions are appropriately sensitive to each other's outputs. It's clear that in cases where the dispositions generate beliefs that are logically inconsistent, the dispositions are not sensitive to each other's outputs. But while logical inconsistency of the outputs suffices to show that the doxastic dispositions are not properly integrated with the subject's cognitive character, it cannot be a necessary condition. If the doxastic disposition *D* generates in me the belief that *p* and another doxastic disposition *D** generates in me the belief [I don't know that *p*] then *D* and *D** are acting in tension with each other, even though *p* and [I don't know that *p*] are not logically inconsistent.

²³ I argue elsewhere (2020) that if knowledge requires employing cognitive abilities that are integrated to our cognitive character, then modal conditions for knowledge which are relativized to such abilities are not hostage to the possible truth of the extended mind thesis.

A tempting way to explain the tension between D and D* is to appeal to the fact that the outputs that they generated cannot amount to knowledge simultaneously. We could then claim that two doxastic dispositions are acting in concert with each other only if it's possible that the outputs amount to knowledge on the condition that both outputs are true. This constraint on cognitive integration is supported by the idea that knowledge is the norm of belief.²⁴ The purpose of our cognitive abilities is to provide a unified picture of the world that amounts to knowledge. If our doxastic dispositions are acting against each other in such a way that achieving this aim is impossible, then at least some of those doxastic dispositions are not integrated with our cognitive character.

However, by saying that two doxastic dispositions can act in concert with each other just in case the outputs they yield could have amounted to knowledge simultaneously threatens to make virtue reliabilism a circular theory of knowledge. While this would probably suit knowledge first virtue reliabilists like Kelp and Miracchi, it's doubtful whether those who aim to provide a reductive virtue-theoretic analysis of knowledge should understand cognitive integration in this way.

But here it's important to note that we need not commit ourselves to the idea that cognitive integration should ultimately be understood in terms of knowledge. Rather, we can only note that when it's in principle impossible that the two outputs could have amounted to knowledge if they were true, then the doxastic dispositions that produced the outputs are not acting in concert. True, we will use our pre-theoretic understanding of knowledge when determining whether a doxastic disposition is integrated to the subject's cognitive character, as does Williamson (2000) when he uses our pre-theoretic understanding of knowledge to determine whether a belief is safe. But this need not make virtue reliabilism a circular theory of knowledge. Virtue reliabilists are still free to unpack the notion of cognitive integration without appealing to knowledge. All we require here is that the way in which virtue reliabilists end up

²⁴ The knowledge norm of belief has been endorsed by Williamson (2000) and Sosa (2011) among many others.

unpacking cognitive integration entails that two doxastic dispositions that are acting in a ‘knowledge-inconsistent way’ are not acting in concert with each other.

Finally, it’s worth keeping in mind that virtue reliabilists relativize cognitive abilities to normal or appropriate conditions and environments (Greco, 2010; Sosa, 2010). This means that cognitive abilities can be lost when moving to environments that are not suitable for the use of those abilities. The fact that one’s doxastic dispositions don’t act in concert in some such conditions and environments doesn’t mean that those doxastic dispositions wouldn’t qualify as cognitive virtues in more suitable environments and conditions, where the doxastic dispositions in question are in the market of being cognitive abilities. This helps to alleviate the pressure to think that INTEGRATION is a too strong condition.

In the next section we examine under what kind of conditions a belief can serve as a defeater.

3. DEFEAT AND JUSTIFICATION

I’ll assume that only those beliefs that have a positive epistemic status can serve as defeaters. I think that this positive epistemic status is justification. Irrational and unjustified beliefs cannot serve to defeat knowledge or justified beliefs. I take this to be the mainstream position among epistemologists²⁵, but it’ll be useful to go through the rationale for this position, since it plays a pivotal role in the next section.

Often some of our beliefs confer justification on our other beliefs. The fact that I know that the drink is laced with hemlock justifies me in believing that the drink is poisonous. In this case

²⁵ See for instance Klein (1976), Lasonen-Aarnio (2010), Pryor (2000), Baker-Hyatt and Benton (2015), Casullo (2018), Pollock (1987), and Alston (2002). Some of these authors understand defeaters in terms of ‘evidence’ or ‘reasons’, rather than in terms of ‘justification’. Since virtue reliabilist theories of knowledge don’t tend to invoke notions such as ‘evidence’ and ‘reasons’, but rather understand them in terms of epistemic competences that are relevant for justification (Sylvan & Sosa, 2018), I prefer to talk about justification here. At any rate, this terminological difference is of little importance. The central idea here is that a belief needs to have some positive epistemic status in order to be a putative defeater. Plantinga (2002) and Bergmann (2006) are notable exceptions from the mainstream position since they think that even irrational beliefs can serve as defeaters. Thanks to Maria Lasonen-Aarnio for discussion.

my knowledge entails the truth of the latter belief. But if I believed out of sheer paranoia that the drink is laced with hemlock, I wouldn't be justified in believing that the drink is poisonous. While the contents of my beliefs in the above cases stand in exactly the same logical relations, my belief that the drink is poisonous isn't justified in the latter case, since there is no justification to be transmitted from my belief that the drink is laced with hemlock. Similarly, if I were to believe out of wishful thinking that England is going to lose the game, I wouldn't thereby be justified in believing that Italy is going to win the game. If justified beliefs could be built on paranoia and wishful thinking living a good epistemic life would be all too easy.

Given that irrational and unjustified beliefs cannot confer positive epistemic statuses on our other beliefs, it would be *prima facie* bizarre if they could render our justified beliefs unjustified. How could they have only this kind of negative epistemic import? Moreover, if irrational and unjustified beliefs can defeat justified beliefs, then they can also serve to *restore* the justificatory status of beliefs (Casullo, 2018).²⁶ This is because a putative defeater *d* can be defeated by yet another putative defeater *d'*, rendering the original belief justified once again (Pollock, 1987). One shouldn't be able to restore the justificatory status of a defeated belief by irrationally believing that the putative defeater doesn't defeat one's original belief. Otherwise irrational and unjustified beliefs can confer justification to our beliefs. Therefore, only justified beliefs can serve as defeaters.

Virtue reliabilists think that a subject S's belief is justified if, and only if it's an exercise of S's cognitive abilities (Greco, 2002, p. 311; 2010; Kelp, 2017, p. 238; Miracchi, 2015, p. 48; Sosa, 1991, p. 189). Given that beliefs need to be justified in order to serve as defeaters, a defeater-belief must be a product of one's cognitive abilities.

4. DEFEAT OF THE VIRTUES?

²⁶ See also Alston (2002, pp. 198-199) for an argument that unjustified beliefs cannot serve to defeat justified beliefs.

So far I've shown that virtue reliabilists are committed to the idea that knowledge arises from exercises of cognitive abilities and that a doxastic disposition can qualify as a cognitive ability only if it's suitably integrated with the cognitive character of the subject. I've also explained that in order for a putative defeater to have potential normative import, it must be the case that the defeater enjoys a positive epistemic standing. I assume that the defeater has to be justified in order to have potential normative import. On virtue reliabilism justified beliefs are exercises of cognitive abilities. Therefore, the defeater belief has to be an exercise of a cognitive ability in order to have potential normative import. Given this, what must virtue reliabilists say about the phenomenon of knowledge defeat?

Consider a paradigmatic case of knowledge defeat like **Red light**:

At t_1 Yen comes to know that the wall in front of her is red via perception in optimal conditions. At t_2 Yen's trusted friend Triss tells her that the wall is illuminated by red light that would have made the wall look red even if it had been of some other colour.

In order for **Red light** to be a potential case of knowledge defeat it must be the case that Yen's belief that the wall is red is a product of her cognitive abilities. Otherwise her belief could not have qualified as knowledge at t_1 . It must also be the case that her belief that [the wall is illuminated by red light that would have made the wall look red whatever its actual colour is] is a product of her cognitive abilities since otherwise the defeater belief wouldn't be justified, and hence wouldn't have any defeating force. Now suppose that Yen dogmatically clings to her belief that the wall is red after forming a justified belief in the defeater. Given that the defeater supports that her original belief doesn't qualify as knowledge, what should virtue reliabilists say about this case?

I think that virtue reliabilists are committed to claiming that the case, when described in this way, is metaphysically impossible. It cannot be the case that both Yen's original belief and her defeater belief are products of her cognitive abilities. Why? Because the doxastic dispositions

that generate these beliefs are clearly acting in tension, rather than in concert with each other. After all, the way in which Yen believes that the wall is red can only constitute knowledge if her defeater belief isn't knowledge and *vice versa*. This is because if Yen knows by visual perception alone that the wall is red at t_1 it cannot be the case that wall is bathed in red light at t_1 , because then the colour that Yen would have seen would be that which the red light cast on the wall and not the redness of the wall. The truth of Yen's perceptual belief would lack an appropriate causal connection to what makes it true. Similarly, Yen cannot know via testimony that the wall was bathed in red light at t_1 if she knew by visual perception alone that the wall is red at t_1 . After all, if the wall was bathed in red light at t_1 the truth of Yen's perceptual belief would have lacked an appropriate casual connection to what makes it true, and hence she couldn't have known by visual perception that the wall is red. So while the contents of the doxastic outputs are not logically inconsistent, the ways in which the beliefs are formed are epistemically inconsistent in that both beliefs could not have constituted knowledge simultaneously.

Recall that INTEGRATION requires that the subject's doxastic dispositions would act in concert with the other doxastic dispositions that make up the subject's cognitive character if it were triggered. In cases where the subject dogmatically clings onto her belief after forming a justified belief in the putative defeater, this counterfactual is false. Importantly, the counterfactual was already false at the moment when the subject formed her original belief, and hence the subject's original belief was not a product of her cognitive abilities, and cannot qualify as knowledge. Therefore, if the defeater belief is justified, and the subject holds onto her original belief after receiving the putative defeater, her original belief never amounted to knowledge to begin with. Since the subject never acquires knowledge in this *first variant* of the case, there is no knowledge defeat.

Alternatively, it could be the case that Yen's defeater belief isn't a product of her cognitive abilities, in which case the belief would be unjustified. But if it's true that unjustified beliefs cannot serve as defeaters Yen doesn't have a defeater for her belief that the wall is red.

Moreover, since Yen's defeater belief isn't a product of her cognitive abilities, the doxastic dispositions that help to constitute her cognitive character are not acting in tension with each other if she holds onto her original belief. Therefore, Yen can know that the wall is red. And since Yen can continue to know in this *second variant* of the case that the wall is red after t_2 there is no knowledge defeat in this variant either.

But suppose that instead of dogmatically holding onto her belief, Yen suspends judgment after having formed a justified belief in the defeater. In this *third variant* Yen's original belief might have amounted to knowledge, since the cognitive abilities that are responsible for her perceptual belief are acting in concert with her cognitive character. In this version Yen is acting in the same way as the subject who cannot fail to see the Müller-Lyer lines as being of different lengths but nevertheless doesn't believe that they are of different lengths after having learned, perhaps by testimony, that the Müller-Lyer lines constitute a known illusion. But while Yen's original belief and her defeater belief might be products of her cognitive abilities in this variant of the case, there is no knowledge defeat in this case either. It's true that she doesn't know that the wall is red after having received the putative defeater, but this is because she doesn't believe that the wall is red after having received it. It's not the defeater that robs her of knowledge; it's her lack of belief. Below is a table that summarizes these different variants.

	Original belief justified	Defeater belief justified	Suspends judgment	Knowledge defeated
Variant 1	No	Yes	No	No
Variant 2	Yes	No	No	No
Variant 3	Yes	Yes	Yes	No

But while knowledge defeat doesn't occur in any of the three variants, it's not impossible for Yen to lose her knowledge while holding onto her original belief. If Yen were to rebase her belief that the wall is red at t_2 , and the doxastic disposition responsible for the rebasing was not a cognitive ability, she would fail to know that the wall is red at t_2 , even though she would still believe that the wall is red. But here it's important to recall that whether Yen rebases her belief at t_2 is a contingent matter. And since it's a contingent matter, Yen doesn't necessarily lose her

knowledge after having acquired the putative defeater. It could also be the case that Yen's cognitive character changes between t_1 and t_2 in such a way that the doxastic disposition that generated the belief that the wall is red no longer counts as a cognitive ability at t_2 .²⁷ In this variant of the case Yen could have known at t_1 that the wall is red, but doesn't know it at t_2 , since the doxastic disposition in charge of retaining the belief doesn't qualify as a cognitive ability at t_2 . But again, it's a contingent matter whether Yen's cognitive character changes between t_1 and t_2 , and hence the fact that knowledge is lost in this variant of the case doesn't suffice to show that Yen's knowledge is defeated.

Recall that knowledge defeat occurs just in case a subject knows that p , gains a putative defeater for her belief that p , and *thereby* loses her knowledge that p *without necessarily losing her belief* that p . The defeatists claim that acquiring the putative defeater suffices *on its own* to defeat one's knowledge. But knowledge defeat doesn't occur in any of the five variants of **Red light** that we just considered. In the first version Yen never knew, in the second one she never gained a defeater, and in the third one she lost knowledge only because she lost the relevant belief. In the last two variants Yen does lose her knowledge without losing the corresponding belief. But this is only because she either (i) starts believing that the wall is red via a method of belief-formation that isn't a cognitive ability, or, (ii) her cognitive character changes in such a way that the way in which she formed her belief originally no longer counts as a cognitive ability. Yen doesn't lose her knowledge in any of these variants solely in virtue of having acquired a putative defeater for her belief.

So knowledge defeat, strictly speaking, is an illusory phenomenon. There are no cases where acquiring a putative defeater for a belief that qualifies as knowledge suffices on its own to defeat the belief's epistemic standing. But while knowledge defeat turns out to be an illusory phenomenon on the sketched account, it's nevertheless true that there are no cases where a

²⁷ As far as I know virtue reliabilists have not discussed whether a subject's cognitive character could change. Nevertheless, it is obvious that a subject can gain and lose cognitive abilities, and hence subjects' cognitive characters can change.

subject knows that p after having acquired a putative defeater for her belief that p . Thus virtue reliabilists are able to explain intuitions of knowledge defeat, without granting that knowledge defeat is a genuine phenomenon. Virtue reliabilism provides an error-theory of our defeat intuitions. It's an error-theory in two senses. First, it claims that in some putative cases of knowledge defeat, knowledge was never had to begin with. Second, the potential loss of knowledge isn't explained in terms of the putative defeater's normative force, but rather via the way in which the subject reacted to her epistemic situation.²⁸ Finally, it's worth noting that this error-theory can explain why suspending judgment is nevertheless epistemically speaking good, even though putative defeaters lack normative force. Suspending judgment is epistemically optimal, because only in those cases where the subject suspends her judgment is it possible that both her original, and her defeater belief, were justified (variant 3 above).²⁹

Here's an objection I've heard against the theory proposed (voiced by Maria Lasonen-Aarnio, among others). Intuitively Yen knows in **Red light** that the wall is red at t_1 even if she would dogmatically cling onto her belief if she later gained a justified belief that is a putative defeater for her original belief (variant 1 above). Yen's dogmatism is a vice of her epistemic character that doesn't stain her belief, the objection goes. I grant the objector that intuitively Yen knows at t_1 that the wall is red. While it might be unintuitive that Yen's dogmatism would preclude her from knowing that the wall is red, virtue reliabilists are committed to this claim. They hold that knowledge and justification can only arise from the exercise of cognitive abilities that are

²⁸ Note that though this error-theory goes against the defeatist orthodoxy, it does fit with our ordinary practice of attributing knowledge. Azzouni (2020, p. 334), aptly notes that "it's not fairly straightforward common sense to think that misleading evidence can undercut knowledge... *We never say...* "I knew p once, but then, because usually trustworthy S lied to me, I stopped knowing it." I would like to thank an anonymous reviewer at *Erkenntnis* for alerting me to Azzouni's work.

²⁹ Neta (2002, pp. 675-676) has provided a contextualist theory of knowledge yields an account of defeat that bears some similarity to the account proposed here, in that according to it acquiring new evidence cannot on its own defeat knowledge. I'd like to thank an anonymous reviewer at *Erkenntnis* for pointing this out.

integrated to one's cognitive character.³⁰ Virtue reliabilists can explain the intuition that Yen knows that the wall is red at t_1 . In all but variant 1 Yen does know that the wall is red at t_1 . It's easy to mix up the variants since information regarding Yen's dogmatic character is revealed only later. Furthermore, variant 1 is, perhaps, the most unnatural way of fleshing out the case. Most people would withdraw their belief if they were presented with a putative defeater. It's plausible that we implicitly assume that Yen is non-dogmatic when originally evaluating whether Yen knows at t_1 . First impressions are hard to shake, especially when it comes to intuitions. That said, those who think that the objection is successfully, are invited to see this paper as offering an argument against virtue reliabilism. My aim was to examine what virtue reliabilists ought to say about defeat given some of their core commitments.

5. OTHER VIRTUE THEORETIC PROPOSALS

I will briefly consider some alternative solutions to the problem of knowledge defeat, put forth by virtue reliabilists. In order to deal with defeaters Greco (2010) adds a subjective justification condition to his analysis of knowledge. According to Greco subject S 's belief that p is subjectively justified "if and only if S 's believing that p is properly motivated; if and only if S 's believing that p results from intellectual dispositions that S manifests when S is motivated to believe the truth" (2010, p. 167). Greco's argument as to why knowledge entails subjective justification is motivated by his take on Aristoteles's virtue ethics. According to him virtuous action requires not only that the action arises from a virtuous character trait, but also that the action is properly motivated by one's virtuous character (Greco, 2010, p. 43).

I won't take issue with the difficult question under what conditions a subject is properly motivated to believe the truth, nor with Greco's motivation to add a subjective justification component to his analysis of knowledge. For the sake of the argument, I'll also grant that in putative cases of knowledge defeat Greco's subjective justification condition isn't satisfied and

³⁰ Hurka (2006), Sylvan (2017) and Lasonen-Aarnio (forthcoming-a) have criticized character-attaching virtue theories, like virtue reliabilism, for requiring that virtuous acts must arise from virtuous character.

that knowledge is hence lost in such cases. I only wish to note that the kind of virtue reliabilism that Greco endorses already has the necessary tools to explain our intuitions of knowledge defeat. Adding a subjective justification condition to the analysis isn't necessary and achieves nothing on this score. To me, adding this condition seems like an extra cost.

Pritchard (2018) has argued that his anti-luck virtue epistemology can account for the phenomenon of knowledge defeat. He claims that a subject who comes to know that [that's a barn] in an area with no barn facades around, loses her knowledge if she sees a sign that says that she is in the barn façade-county. He writes that "the safety of her cognitive success is now *in spite of* her manifestation of relevant cognitive agency, rather than being to any significant degree because of it" (2018, p. 3075). Assuming that her belief that [that's a barn] is a product of her cognitive abilities, I fail to see why the subject's safe cognitive success wouldn't be to a significant degree because of the exercise of her cognitive abilities. After all, the fact that the subject trusts her perception seems to explain precisely why she continues to have a safe belief. Pritchard needs to tell us more about why the subject's safe cognitive success isn't attributable to her cognitive agency in cases like this, if his explanation of knowledge defeat is to succeed. Moreover, if I am correct, Pritchard already has the necessary tools to accommodate our intuitions of knowledge defeat.

As far as I know, Sosa has not addressed the problem of knowledge defeat in print. However, given his distinction between animal and reflective knowledge, he could perhaps adopt the following view.³¹ In putative cases of knowledge defeat one's original belief retains its aptness, and hence it amounts to animal knowledge. However, once the defeater is introduced, the subject can no longer aptly take her belief to be apt, which is what reflective knowledge would require (Sosa, 2011, p. ch. 1). Reflective knowledge requires that one competently assesses the risk of forming a false belief to be low enough, and arguably, one cannot have competently assessed the risk to be low enough if one has a defeater for one's belief. Therefore, defeaters would destroy reflective knowledge, but leave animal knowledge intact. This error-theoretic

³¹ I'd like to thank Markus Lammenranta for suggesting this error-theory on Sosa's behalf.

account of knowledge defeat rests on Sosa's distinction between animal and reflective knowledge. Since the error-theory that I gave is derivable from the core tenets of virtue reliabilism, it's simpler than the possible account that Sosa's more complicated framework could yield. Moreover, if I am correct, Sosa has the resources to accommodate our intuitions of knowledge defeat without resorting to his distinction between animal and reflective knowledge.

To wrap up, the error-theory that I have presented is preferable to extant virtue reliabilist accounts of defeat since it is simpler than those accounts and stems from the core ideas of virtue reliabilism. Virtue reliabilists need not add bells and whistles to explain defeatist intuitions.

6. CONCLUSIONS

I argued that virtue reliabilism is able to explain our defeat intuitions via a new kind of error-theory that falls directly out of the core tenets of virtue reliabilism. According to the error-theory, in paradigmatic cases of knowledge defeat where the subject holds onto her belief, the subject never knew to begin with. In cases where the subject suspends her judgment upon receiving the defeater she might have originally known, but doesn't anymore, since she lacks the relevant belief. In neither case is knowledge lost solely in virtue of the fact that the subject acquired a defeater, and hence knowledge defeat is an illusory phenomenon. Nevertheless, the defeatists are right in claiming that there are no cases where a subject retains her knowledge of p after having acquired a defeater for p .

Acknowledgments: I would like to thank Maria Lasonen-Aarnio, Niall Paterson, Giada Fratantonio, Daniel Drucker, Kurt Sylvan, Clayton Littlejohn, Kate Nolfi, Markus Lammenranta, Matti Eklund, Carl Montan, two anonymous reviewers at *Erkenntnis* and audiences at Universities of Lisbon and Helsinki for helpful comments. Special thanks to Maria and Niall for excellent written comments and extensive discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 758539.

References:

- Alston, W. (2002). Plantinga, Naturalism, and Defeat. In J. Beilby (Ed.), *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism* (pp. 176-203). Ithaca: Cornell University Press.
- Azzouni, J. (2020). *Attributing Knowledge: What It Means to Know Something*. New York: Oxford University Press.
- Baker-Hytch, M., & Benton, M. A. (2015). Defeatism Defeated. *Philosophical Perspectives*, 29(1), 40-66. doi:doi:10.1111/phpe.12056
- Beddor, B. (2015). Process Reliabilism's Troubles With Defeat. *The Philosophical Quarterly*, 65(259), 145-159. doi:10.1093/pq/pqu075 %J The Philosophical Quarterly
- Beddor, B., & Pavese, C. (2020). Modal Virtue Epistemology. *Philosophy and Phenomenological Research*, 101(1), 61-79. doi:<https://doi.org/10.1111/phpr.12562>
- Bergmann, M. (2006). *Justification without Awareness*. New York: Oxford University Press.
- Bonjour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy*, 5, 53-73.
- Breyer, D., & Greco, J. (2008). Cognitive Integration and the Ownership of Belief: Response to Bernecker. *Philosophy and Phenomenological Research*, 76(1), 173-184.
- Brown, J. (2018). *Fallibilism: Evidence and Knowledge*. Oxford: Oxford University Press.
- Casullo, A. (2018). Pollock and Sturgeon on defeaters. *Synthese*, 195(7), 2897-2906. doi:10.1007/s11229-016-1073-5
- Comesana, J. (2005). Unsafe knowledge. *Synthese*, 146(3), 395-404. doi:10.1007/s11229-004-6213-7
- Goldman, A. (1979). What is Justified Belief? In G. S. Pappas (Ed.), *Justification and Knowledge* (pp. 1-25). Dordrecht: Reidel.
- Goldman, A. (1986). *Epistemology and Cognition*. Cambridge: Harvard University Press.
- Goodman, J., & Salow, B. (2018). Taking a chance on KK. *Philosophical Studies*, 175(1), 183-196. doi:10.1007/s11098-017-0861-1
- Greco, J. (1999). Agent reliabilism (Hume, skepticism, truth of empirical beliefs, epistemology). *Nous*, 273-296.
- Greco, J. (2002). Virtues in Epistemology. In P. Moser (Ed.), *Oxford Handbook of Epistemology* (pp. 287--315): Oxford University Press.
- Greco, J. (2010). *Achieving Knowledge: A Virtue Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- Greco, J. (2012). A (Different) Virtue Epistemology. *Philosophy and Phenomenological Research*, 85(1), 1-26. doi:10.1111/j.1933-1592.2011.00567.x
- Hirvelä, J. (2017). Is it Safe to Disagree? *Ratio*, 30, 305-321.
- Hirvelä, J. (2018). On Virtue, Credit and Safety. *Grazer Philosophische Studien*, 95(1), 98-120.
- Hirvelä, J. (2019a). Global safety: how to deal with necessary truths. *Synthese*, 196(3), 1167-1186. doi:10.1007/s11229-017-1511-z
- Hirvelä, J. (2019b). Knowing Without Having The Competence to Do So. 8(2), 110-118. doi:10.1002/tht3.411
- Hirvelä, J. (2020). How to stay safe while extending the mind. *Synthese*, 197(9), 4065-4081. doi:10.1007/s11229-018-01920-0
- Hirvelä, J., & Lasonen-Aarnio, M. (forthcoming). The Cake Theory of Credit. *Philosophical Topics*.
- Hurka, T. (2006). Virtuous act, virtuous dispositions. *Analysis*, 66(289), 69-76. doi:doi:10.1111/j.1467-8284.2006.00591.x
- Kelp, C. (2013). Knowledge: The Safe-Apt View. *Australasian Journal of Philosophy*, 91(2), 265-278. doi:10.1080/00048402.2012.673726
- Kelp, C. (2017). Knowledge First Virtue Epistemology. In A. Carter, E. Gordon, & B. Jarvis (Eds.), *Knowledge First: Approaches in Epistemology and Mind* (pp. 224-245). Oxford: Oxford University Press.

- Klein, P. D. (1976). Knowledge, Causality, and Defeasibility. *The Journal of Philosophy*, 73(20), 792-812. doi:10.2307/2025680
- Lackey, J. (1999). Testimonial knowledge and transmission (Considerations on the epistemology of testimony and cognition). *Philosophical Quarterly*, 49(197), 471-490. doi:10.1111/1467-9213.00154
- Lackey, J. (2007). Why We Don't Deserve Credit for Everything We Know. *Synthese*, 158(3), 345-361. doi:10.1007/s11229-006-9044-x
- Lackey, J. (2009). Knowledge and credit. *Philosophical Studies*, 142(1), 27-42. doi:10.1007/s11098-008-9304-3
- Lasonen-Aarnio, M. (2010). Unreasonable Knowledge. *Philosophical Perspectives*, 24(1), 1-21.
- Lasonen-Aarnio, M. (2014). Higher-Order Evidence and the Limits of Defeat. 88(2), 314-345. doi:10.1111/phpr.12090
- Lasonen-Aarnio, M. (forthcoming-a). Competent Failure and Victims of Deceit. In F. Dorsch & J. Dutant (Eds.), *The New Evil Demon Problem*. Oxford: Oxford University Press.
- Lasonen-Aarnio, M. (forthcoming-b). Dispositional Evaluations and Defeat. In J. Brown & M. Simion (Eds.), *Reasons, Justification and Defeat*. Oxford: Oxford University Press.
- Littlejohn, C. (2014). Fake Barns and False Dilemmas. *Episteme—a Journal of Individual and Social Epistemology*, 11(4), 369-389. doi:10.1017/epi.2014.24
- Lyons, J. (2009). *Perception and Basic Beliefs*. New York: Oxford University Press.
- Lyons, J. (2016). Goldman on Evidence and Reliability. In B. McLaughlin & H. Kornblith (Eds.), *Goldman and His Critics*. Oxford: Blackwell.
- Miracchi, L. (2015). Competence to Know. *Philosophical Studies*, 172, 29-56.
- Neta, R. (2002). S Knows That P. *Nous*, 36(4), 663-681.
- Neta, R., & Rohrbaugh, G. (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly*, 85(4), 396-406. doi:10.1111/j.1468-0114.2004.00207.x
- Nozick, R. (1981). *Philosophical Explanations*. Oxford: Oxford University Press.
- O'Callaghan, C. (2012). Perception and Multimodality. In E. Margolis, R. Samuels, & S. Stich (Eds.), *Oxford Handbook of Philosophy of Cognitive Science* (pp. 92-177). Oxford: Oxford University Press.
- Palermos, S. O. (2014). Knowledge and cognitive integration. *Synthese*, 191(8), 1931-1951. doi:10.1007/s11229-013-0383-0
- Paterson, N. (forthcoming). Safety and Necessity. *Erkenntnis*, 1-17.
- Plantinga, A. (1993). *Warrant: The Current Debate*. Oxford: Oxford University Press.
- Plantinga, A. (2002). Reply to Beliby's Cohorts. In J. Beilby (Ed.), *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism* (pp. 204-275). Ithaca: Cornell University Press.
- Pollock, J. (1987). Defeasible Reasons. *Cognitive Science*, 11, 481-518.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford: Oxford University Press.
- Pritchard, D. (2010). Cognitive ability and the extended cognition thesis. *Synthese*, 175, 133-151. doi:10.1007/s11229-010-9738-y
- Pritchard, D. (2012). Anti-Luck Virtue Epistemology. *Journal of Philosophy*, 109(3), 247-279.
- Pritchard, D. (2018). Anti-luck virtue epistemology and epistemic defeat. *Synthese*, 195(7), 3065-3077. doi:10.1007/s11229-016-1074-4
- Pryor, J. (2000). The Skeptic and the Dogmatist. *Nous*, 34(4), 517-549. doi:10.1111/0029-4624.00277
- Sainsbury, R. (1997). Easy Possibilities. *Philosophy and Phenomenological Research*, 57, 907-919.
- Shoemaker, D. (2015). *Responsibility from the Margins*. New York: Oxford University Press.
- Silva, P. (2017). Knowing How to Put Knowledge First in the Theory of Justification. *Episteme*, 14(4), 393-412.
- Sosa, E. (1991). *Knowledge in Perspective - Selected Essays in Epistemology*. Cambridge: Cambridge University Press.
- Sosa, E. (1999). How to Defeat Opposition to Moore. *Philosophical Perspectives*, 13, 141-154.

- Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge: Volume I*. Oxford: Oxford University Press.
- Sosa, E. (2009). *Reflective Knowledge: Apt belief and Reflective Knowledge, Volume II*. Oxford: Oxford University Press.
- Sosa, E. (2010). How Competence Matters in Epistemology. *Philosophical Perspectives*, 24(1), 465-475. doi:10.1111/j.1520-8583.2010.00200.x
- Sosa, E. (2011). *Knowing Full Well*. Princeton: Princeton University Press.
- Sosa, E. (2015). *Judgment and Agency*. Oxford: Oxford University Press.
- Sylvan, K. (2017). Responsibilism out of Character. In *Epistemic Situationism*. Oxford: Oxford University Press.
- Sylvan, K., & Sosa, E. (2018). The Place of Reasons in Epistemology. In D. Star (Ed.), *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press.
- Thomson, J. (1997). The Right and the Good. *The Journal of Philosophy*, 94(6), 273-298. doi:10.2307/2564542
- Turri, J. (2011). Manifest Failure: The Gettier Problem Solved. *Philosophers Imprint*, 11(8), 1-11.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
- Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.