Knowledge, Despite Evidence to the Contrary*

Plato claimed, in the *Meno* (97a—98c), that knowledge is more valuable than true belief in virtue of it being more *stable* than true belief. Knowledge, Plato argued, is a more reliable guide to action than true belief: the traveler who *knows* the way to Larissa is less likely to make a wrong turn (or to give wrong directions) than the traveler who has a mere *true belief* of the way to Larissa, because known truths are — and merely believed truths aren't — *tethered* to one's mind[1] by an explanation of why they are true (*aitias logismos*).[2] In other places (e.g., in *Republic* 412e ff.) Plato also seems to have claimed that a truth, once tethered, is not easily (or perhaps ever) *un*tethered by counterevidence.[3] The main

---

*I wrote this paper to honor Peter Klein, the best philosopher I have ever met, and one of the best people I have ever met. Peter was my thesis advisor at Rutgers; he is also a mentor, and the best friend my work has ever had. Peter made me believe I could (sometimes) do philosophy, and that philosophy was something worth doing. During the eight years I was at Rutgers, Peter and I met once or twice every month during the academic year to discuss my work. We discussed every issue in this paper (and more, much more) innumerable times over that period. We didn't always agree, but Peter always taught me something valuable. I cherish those lessons and I will be forever grateful to Peter for having given them to me.

[1]Plato's discussion of the difference in value between knowledge and true belief famously relates knowledge and true belief to Daedalus statues, which, if not tethered, were said to run out and escape their base. A true belief, argued Plato, is like an untethered Daedalus statue — beautiful but prone to escaping its owner. An instance of knowledge, on the other hand, is like a Daedalus statue that *is* tethered — beautiful and not likely to escape its owner.

[2]I am aware of the pitfalls of trying to translate 'aitias logismos,' but I will not discuss them here. For an excellent discussion of some of the main exegetical issues surrounding the translation this expression, see Fine (2004).

[3]For example, in *Republic* 534b Plato seems to argue that knowledge is immune to counterevidence, while mere true belief is not, because the latter, but not the former, may be essentially based on false beliefs. Because mere

goal of this paper is to understand when, if ever, knowledge is, to use Plato's expression, *un*tethered by evidence suggesting that what one knows is false.

The idea that knowledge is *not* defeated by counterevidence enjoys some popularity among contemporary philosophers. For example, Norman Malcolm (1952, p.185—186) wondered whether there is any evidence that should be allowed to lower his confidence in a proposition he says he knows; that there is an ink—bottle in front of him. Malcolm concluded his rumination thus (emphases are all his):

> ... I should regard nothing as evidence that there is no ink—bottle here now... [This] describes my *present* attitude towards the statement that here is an ink—bottle. ... My present attitude toward that statement is radically different from my present attitude toward ... other statements (e.g., that I have a heart). I do *now* admit that certain future occurrences would disprove the latter. Whereas no imaginable future occurrence would be considered by me *now* as proving that there is not an ink—bottle here. These remarks are not meant to be autobiographical. They are meant to throw light on the common concepts of evidence, proof, and disproof. (Malcolm, 1952, p.181—2)

According to Malcolm, the case highlights the *strong sense* of 'know.' He contrasts a strong sense of 'know' with a *weak sense* of that word. Central to Malcolm's distinction is the idea that the weak sense of 'know' allows for the possibility of refutation, while the strong sense does not. For example, when one says that one *knows* that 92 times 16 is 1,472, but one is not sure this is the case (say, because one did the calculation in one's head), 'know' is being used in its weak sense; when one says that one *knows* that 2 plus 2 is 4, one is sure this

---

true beliefs can depend essentially on false beliefs, they may be undermined by new truths; this is not the case with knowledge, which depends exclusively on what is true. See (Fine, 2004, sec. viii) for a discussion of this point.

is the case and 'know' is being used in its strong sense.[4] I will make a similar suggestion about knowledge below. However, unlike Malcolm, who focused on the ordinary use of 'know,' my focus will be on the concept of knowledge.

I

We can easily reproduce cases that seemingly show that knowledge may be retained in the face of counterevidence. Consider:

> *Headache*
> Liz has a throbbing headache. On her way to the medicine cabinet she bumps into her father, a psychiatrist, who proceeds to tell Liz that she does not have a headache. He argues that Liz's subconscious mind is playing tricks on her in order to allow her to deal with the sudden death of her dog, Fifi.[5]

Although the testimony of Liz's father provides her with counterevidence to her knowledge that she has a headache, that does not defeat the justification she

---

[4]Malcolm is making another distinction with his case. Although he is primarily interested in the *synchronic* features of the case (i.e., he is interested in what it is *now* rational for him to think about possible counterevidence to his knowledge that there is an ink—bottle in front of him), he acknowledges that one could also take an interest in the *diachronic* features of the case (i.e., one could be interested in what Malcolm *would* do, in the future, if he were confronted with evidence suggesting there was no ink—bottle in front of him). Malcolm says that he is not talking about the diachronic features of his situation, and he concedes that he cannot confidently predict now how he would behave in the future were he to encounter evidence suggesting that there is no ink—bottle in front of him. For all he knows, says Malcolm, he would 'become mad' or 'fall into a swoon' if someone he trusts were to say that ink—bottle hallucinations are quite common for people in his circumstances. See (Kripke, 2011) and (Borges, 2015) for a discussion of how heeding this distinction may prevent one from misunderstanding the charge of dogmatism that is usually leveled against strong views of knowledge such as the one discussed by Malcolm.

[5]I don't mean this case to presuppose that headaches are *luminous* states (i.e., states that are such that, if one is in one of them, then one is in a position to know one is in them). Maybe headaches are luminous; maybe they aren't. The claim that they are is irrelevant for my case. Clearly, we are *sometimes* in a position to know we have a headache; I mean this case to be about *one of those times*. I mention this because Timothy Williamson (2000) has persuasively argued that there are no luminous states. His argument does not preclude the claim that one is *sometimes* in a position to know one is in a certain mental state, though. But see Greenough and Pritchard (2009) for criticism of Williamson's argument against luminosity.

has for believing she has a headache. The case illustrates Malcolm's point that (strong) knowledge is not defeated by counterevidence. Here's another case that seems to make the same point:

> *Weather*
> Liz is looking at a sunny afternoon through the window in her study while listening to a DVRed Weather Channel forecast from last night. The forecaster said, then, and with a fair amount of certainty, that it would be raining right now (i.e., at the time Liz is listening to the recording).

The fact that the forecaster said it would be raining at the time Liz is watching the recording amounts to counterevidence to Liz's knowledge that it is sunny outside. Moreover, she knows that the forecaster said that it would be raining by the time she is listening to the recording. This does not seem to change the fact that Liz is justified in believing that it is sunny outside, and, since it is true that it is sunny outside, she knows that it is.[6] In short, Liz seems entitled not to change the confidence she places in the proposition that it is sunny outside even though she encountered counterevidence for the claim that it is.

Like Plato and Malcolm, I also feel attracted to the claim that knowledge is sometimes justifiably retained in the face of counterevidence to what one knows — *Headache* and *Weather* tell me this much.

II

In contrast with the Plato—Malcolm tradition that takes knowledge to be indefeasible, a different (and at least equally as influential) tradition takes knowledge

---

[6] Of course, I am also supposing that there is no funny business of the Gettier type going on here. The same is true of all the other cases in this paper.

to be defeasible. Gilbert Harman, Timothy Williamson, and many others have given voice to this tradition.[7] For instance, here is Williamson making quite succinctly the point that knowledge is defeasible :

> The assumption [that knowledge can be added but not subtracted over time] is obviously false in practice, because we sometimes forget. But even if the model is applied to elephants, idealized subjects who never forget, the assumption that [knowledge] cannot be lost is implausible. On any reasonable theory of [knowledge], an empirical proposition which now counts as [knowledge] can subsequently lose its status as [knowledge] without any forgetting, if future evidence casts sufficient doubt on it.[8]

This point is usually backed up by examples such as this:[9]

Marbles
Mary puts a red and a black marble in a bag. She shakes the bag well and proceeds to make draws with replacement while carefully taking notes of the result of each draws. Each of Mary's first one thousand draws produce a red marble.[10]

---

[7] For example, Jaakko Hintikka (1962, p.20), claims that 'whoever says "I know that p" proposes to disregard the possibility that further information would lead him to deny that p.' Peter Klein (2017), and Roy Sorensen (2012), also endorse the view that knowledge is defeasible. I will discuss Peter Klein's views below.

[8] (Williamson, 2000, p.206).

[9] Gibert Harman (1973, p.192) does just that in the following passage:

> That undermining evidence ... is relevant ... to the acquisition of knowledge [and] also to its maintenance is clear from the ... example concerning Tom and the library detective. The detective knows that he saw Tom steal the book and so he testifies to the Judicial Council. After he leaves the hearing, Tom's mother fabricates her story about Tom's twin brother. Her lying testimony convinces the Judicial Council but is unknown to the detective back at his post in the library. [O]nce Tom's mother has ... testified, it is no longer true that the detective knows that Tom stole the book.

[10] *Marbles* is a variation of a case discussed by Williamson (2000, p.205).

The natural reaction to *Marbles* seems to be something like this: although Mary knows, at first, that there is a red and a black marble in the bag, after drawing a red marble one thousand times Mary ceases to know that this is the case, even though, by stipulation, this is true. Her evidential situation changes as the case progresses, however. She starts the case with excellent evidence for the claim that the bag has both a red and a black marble, but each time Mary draws a red marble from the bag, she acquires a small piece of counterevidence to the claim that the bag has both a red and a black marble. By the one—thousandth time she draws a red marble, the small pieces of counterevidence have added up to more counterevidence than her knowledge—level justification could bear. Mary's knowledge that there is a red *and* a black marble in the bag is thus defeated; its justification dies a death by one thousand counterevidential cuts.[11] Knowledge, *Marbles* suggests, is not completely tethered, or fixed; it is in fact somewhat *loose* and may, as it were, be pushed out of one's mind. The line of reasoning in this reaction to *Marbles* is intuitive, and apparently in tension with the idea that knowledge is tethered or stable. We will get back to this tension below.

Here one might feel inclined to complain and say that the case is too implausible to support the epistemic point I am making. After all, what is the chance that someone will draw the same red marble one thousand times in a row?![12] In order to assuage these fears, we may consider an analogous case that takes care of this issue.

Marbles*
Things are as before except for the fact that Mary's draws are red or

---

[11] *N.B.* I am not making a claim about *exactly* what impact each draw has on her justification; the impact of each draw is plausibly really small. The point, rather, is that *the sum* of all one thousand draws intuitively does have a major impact on her justification. Thanks to Cherie Braden for discussion here.

[12] Answer: small — 0.5^{1000}.

black more or less at the rate one would expect (i.e., more or less half of the time). After doing this for a couple dozen times, Mary's reliable friend, Larry, walks into the room and says, 'You probably drank tainted water. Two in every three bottles of water in the kitchen have been tainted with a hallucinogenic substance that causes people to hallucinate the wrong color of objects — they think non—red objects are red! You have been drinking from a bottle you got from the kitchen, right? So, it's likely that you drank the tainted water.' Suppose that Larry's testimony is true and that Mary has no reason to doubt him — quite the opposite, Mary has all the reason to trust him, since he is a good reliable friend and Mary knows that. Suppose also that Mary *did not* drink from the tainted water.

Larry's testimony amounts to counterevidence to Mary's knowledge that there is a red and a black marble in the bag. What is more, although we are supposing it to be true that there is a red and a black marble in the bag, it seems that Larry's testimony defeats the justification Mary had for being confident that there is a red and black marble in the bag. One way to put this is to say that Larry's testimony made it implausible for Mary to neglect the possibility that she drank the tainted water and that she might be hallucinating having put a red marble in the bag. It seems that this is enough to lower the justification Mary had for this claim below the threshold required for knowledge.

While the Plato—Malcolm tradition and cases such as *Weather* and *Headache* suggest that counterevidence *cannot* lower one's justification below the threshold required for knowledge, the Harman—Williamson tradition and cases such as *Marbles/Marbles** suggests the exact opposite.

III

The clash between these two views of knowledge is captured in the following two claims:

(C1) Knowledge is never defeated by counterevidence.

(C2) Knowledge is sometimes (but not always) defeated by counterevidence.

A word of warning is warranted before we progress. I do not mean for the discussion so far to be taken as 'conclusive evidence' or even as 'strong evidence' for (C1) and (C2). Rather, the discussion so far illustrates what epistemologists who accept (C1) and (C2) have in mind when they argue for those claims. The Plato—Malcolm tradition seems to think that the indefeasibility of knowledge is what explains (at least in part) why knowledge is more valuable than true belief. The Harman—Williamson tradition, on the other hand, seems to believe that the defeasibility of knowledge is a consequence of subjects taking evidence seriously. In a sense, the first tradition emphasizes the *difference in quality* between the grounds for knowledge and grounds for mere true belief, while the latter tradition emphasizes what those grounds have in common.

(C1) and (C2) cannot both be true, of course. However, both claims can be false at the same time: if knowledge is *always* defeated by counterevidence, then it is false that it is never defeated or that it is sometimes but not always so defeated. Although there are costs and benefits associated with accepting one of those claims and rejecting the other, I take it that *Weather* and *Headache* show quite conclusively that the idea that knowledge is always defeated by counterevidence is not very plausible. But, more generally, if one's epistemology allows for the *partial defeat* of evidential support, then knowledge is not always defeated by counterevidence, even if evidential support *is* always so defeated. Let me explain.

Since what gets defeated by counterevidence is the epistemic support one's belief receives from one's evidence (i.e., one's justification), if one can know something even if the probability of the proposition believed given one's evidence is lower than 1, then counterevidence will not always defeat knowledge: if S can know that p when the probability that p is true given S's evidence is, say, .95, then the probability that p is true conditional on S's evidence may decrease from, say, .98 to .96 in virtue of some counterevidence without S ceasing to know that p.[13]

There are at least a couple of costs associated with accepting (C1). One of them is having to explain away the intuition that Mary's knowledge in *Marbles/Marbles** is destroyed by counterevidence.[14] The clear benefit of accepting (C1), on the other hand, is arguably *methodological.* (C1) is simple (i.e., there is no need to find a principled way to distinguish between situations in which knowledge is defeated by counterevidence from situations in which knowledge is *not* defeated by counterevidence) and easily falsifiable (i.e., one case in which knowledge is defeated by counterevidence is sufficient to falsify C1). (C1) also fits nicely with the Platonic idea that knowledge is stable in a way that true belief is not, for, according to (C1), knowledge is not only stable but it simply cannot be moved by evidence to the contrary.

As for (C2), the obvious cost associated with this view of knowledge is one of providing a principled way to distinguish situations in which counterevidence defeats knowledge from situations in which it does not. We could get rid of

---

[13] Be that as it may, all cases discussed in this paper deal with categorical, rather than partial, defeat. That is, with respect to all cases discussed in this paper, I am supposing that the potential defeating effect of counterevidence is *always* such that it lowers the subject's degree of justification below the threshold required for knowledge.

[14] There is another, less obvious, potential cost of accepting (C1). That is the danger of dogmatism and the rejection of an intuitive notion of intellectual humility: if knowledge is indefeasible, why should one pay attention to counterevidence to what one knows? Isn't all counterevidence to what one knows misleading evidence? This problem for views that accept (C1) was put forward by Saul Kripke in a talk in the early 1970s. (Harman, 1973) made a similar problem popular. (Kripke, 2011) includes Kripke's talk as well as an appendix explaining the difference between his version of the problem and Harman's. Since I dealt with this issue in (Borges, 2015), I will not discuss it here again.

this cost by emending (C2) and suggesting that counterevidence *always* destroys knowledge, but that emendation would, I take it, make (C2) way too implausible — as *Headache* and *Weather* suggest. On the other hand, the benefit of accepting (C2) does not seem to be methodological simplicity, but something like *fitness*, since it seems to fit our intuitive judgment about all the cases we discussed so far — it takes those cases as data to be *explained* rather than to be explained *away*) and, if correct, accounts for the intuitive judgment that knowledge is not defeated in *Weather* and *Headache*, and the judgment that knowledge is defeated in *Marbles/Marbles**.

Before we move forward, it is useful to distinguish between two different ideas we might want to convey when we say that someone's knowledge is defeated by counterevidence. On the one hand, when we say that S's knowledge that p is defeated by some counterevidence, e, we may mean to say that S's learning that e is the case *caused her to stop believing* that p. On the other hand, when we say that S's knowledge that p is defeated by some counterevidence, e, we may mean to say that S's learning that e is the case *lowers S's degree of justification for believing that p below the threshold required for knowledge*. One can lose knowledge because of counterevidence in either of those two ways: by ceasing to believe or by having one's justification lowered below the threshold required for knowledge.[15] What is more, although those two ways in which knowledge

---

[15] For simplicity's sake, in drawing this distinction I am presupposing that knowing that p entails believing that p. I think that even if one rejected this presupposition one could still draw the relevant distinction, although doing so would be more complicated. What I mean is that the distinction is compatible with the view, sometimes exposed by knowledge firsters, that *knowledge* cannot be analyzed in terms of *belief*, *truth*, etc. The reason why this view is compatible with our distinction is that one may hold it and still think that whenever someone knows that something is the case that person also believes that it is the case; that is, one may consistently hold the view that *knowledge* cannot be analyzed in terms of *belief* while, at the same time, argue that knowing that p is always accompanied by believing that p (or, alternatively, that the latter state is always a by—product of the former state). If one had this kind of knowledge first view, one could consistently hold the distinction between what I will call below 'psychological' and 'epistemic defeat.' (Incidentally, (Fine, 2004) discusses a version of the knowledge first type of view I am sketching here and suggests one might attribute it to Plato.) I do not have the space to defend these claims here, however. I will come back to them some other day.

may be defeated are clearly related (e.g., an awareness of the fact that counterevidence lowered one's justification below the threshold required for knowledge may cause one to stop believing), it is important to keep them apart when assessing the plausibility of (C1) and (C2). In particular, one may distinguish two different versions of each claim:

(C1^b) Knowledge is never defeated by counterevidence that causes one to stop believing.

(C1^j) Knowledge is never defeated by counterevidence that lowers one's degree of justification below the threshold required for knowledge.

(C2^b) Knowledge is sometimes (but not always) defeated by counterevidence that causes one to stop believing.

(C2^j) Knowledge is sometimes (but not always) defeated by counterevidence that lowers one's degree of justification below the threshold required for knowledge.

Once we distinguish between these different versions of (C1) and (C2), we make room for a more nuanced approach to cases like the ones we have been discussing. For example, one may now argue that, although Mary's justification for believing that there is a black and a red marble in the bag is not lowered by her coming to know that she drew a red marble one thousand times, this counterevidence may well cause her to lose her belief in the proposition that there is a red and a black marble in the bag. This would involve accepting both (C1^j) and (C2^b). Call this type of view a *strong view of knowledge*. Alternatively, one may insist that the counterevidence in *Marbles/Marbles** does lower Mary's justification for believing that there is a red and a black marble in the bag below the threshold required for her to know that fact; she also loses her belief in that proposition,

but the epistemically significant event is that her justification gets defeated. This approach to the case would involve accepting both $(C2^j)$ and $(C2^b)$. Call this type of view a *weak view of knowledge*.

Epistemologists who endorse strong views of knowledge believe that knowledge defeat is exclusively a *psychological* phenomenon in the sense that counterevidence to what one knows leads one to lose (at most) one's belief; they reject the claim that knowledge defeat is an *epistemic* phenomenon, however. That is, they reject the view that sometimes counterevidence to what one knows lowers one's justification below the threshold required for knowledge.[16]

Strong views of knowledge elegantly explain what is going on in our four cases: although the counterevidence can never *epistemically* defeat one's knowledge, subjects in cases such as *Marbles/Marbles** have their knowledge *psychologically* defeated. The challenge for strong views is to explain away the intuition that Mary's knowledge in *Marbles/Marbles** suffers epistemic defeat, not only psychological defeat. Following Maria Lasonen-Aarnio (2010), I will call the intuition that the knowledge subjects have in cases such as *Marbles/Marbles** is epistemically defeated, *defeat intuition*. Let us explore this point a little further, for it will become decisive below when we assess a recent version of the strong view of knowledge.

The defeat intuition emerging in relation to cases such as *Marbles/Marbles** amounts to an intuition about *how justified* the target proposition is for the subject given her new total evidence (i.e., once she updates her degree of belief in light of the new evidence she acquires). This means that a theory will do justice to this intuition only if it tells a story about *justification* (i.e., that which accounts for a good chunk of the epistemic difference between true belief and knowledge), rather than if it tells a story about some other — albeit related —

---

[16] For versions of the strong view of knowledge, see, among others, (Dretske, 1971), (Dretske, 1981), (Lasonen-Aarnio, 2010), and (Baker-Hytch & Benton, 2015).

epistemic desiderata.[17]

In section five I will argue that a recent version of the strong view of knowledge is unable to account for the defeat intuition in a way that does justice to the story connecting justification and defeat. I will then offer an alternative account of the defeat intuition that does justice to this story.

According to epistemologists who endorse the weak view of knowledge, however, knowledge is vulnerable to both epistemic and psychological defeat. This type of view also offers an explanation of what is going on in our cases: while Mary's knowledge in *Marbles/Marbles** suffers epistemic and psychological defeat, the knowledge of subjects in *Weather* and *Headache* suffers neither. Weak views of knowledge have a hard time giving a principled account of when counterevidence is strong enough to lower one's justification below the relevant threshold for knowledge. This is an important challenge for a few reasons but no less because we seem to have a better intuitive grip on whether a subject knows than we have on whether a subject has a certain degree of epistemic justification and whether the justification she has is affected by counterevidence in a particular way. Knowledge—first versions of the weak view may have an advantage over belief—first versions of the weak view, for knowledge—first versions of the view pressupose that we have a better grip on knowledge than on justification.

IV

In this section we look at Peter Klein's version of the weak view of knowledge — the defeasibility theory of knowledge. Peter has defended a version of the defeasibility theory of knowledge for over forty—five years now.[18] His account

---

[17] This is true not only of more traditional versions of the weak view of knowledge (i.e., weak views that take knowledge to be analyzed in part by justification), but also of knowledge—first views such as the one in (Williamson, 2000, ch.10), which says that knowledge is not analyzable but is nonetheless susceptible to epistemic defeat.

[18] See, among many others, (Klein, 1971), (Klein, 1981), (Klein, 2008), and (Klein, 2017).

accepts the defeat intuition and is designed to accommodate it. Consider:

(Defeasibility) S knows that p iff (1) p; (2) S believes that p; (3) S is justified in believing p; (4) there is no truth, d, such that the conjunction of d and S's justification, j, fails to justify S in believing that p.

Condition (4) is a *no—defeat* condition in line with (C2^j) above.[19] According to (4), certain truths (i.e., *defeaters*) are such that they lower one's justification below the threshold required for knowledge. So, according to Defeasibility, Mary's knowledge that there is a red and a black marble in the bag is defeated in this way by the truth that Mary drew a red marble one thousand times.

Defeasibility is a fine exemplar of the weak view of knowledge, with seemingly a lot going for it. For one, the view seems to provide a neat account of cases where one has a justified true belief but no knowledge; for example, in the Gettier cases: in those cases there is a truth that is such that the conjunction of the subject's justification with this truth fails to justify the subject in believing the target proposition. This take on Gettier cases also allows Peter to account for *epistemic luck*: it is a matter of luck that the gettiered subject ended up with a justified true belief, for she could have easily believed the defeater and failed to know.

There are clouds on the horizon, however. Recently Peter focused on accommodating the possibility of inferential knowledge that is essentially based on a false premise (a 'useful falsehood,' as he calls it). As he points out, given condition (4) the problem for Defeasibility is that the negation of the false premise seems to be a defeater for the subject's justification. Here's one such case (adapted from (Warfield, 2005)):

Handouts

---

[19] As far as I know Peter also accepts (C2^b), but I will not deal with that aspect of his view here.

Ted carefully counts the number of people present at his talk and reasons: 'There are 53 people at my talk; therefore my 100 handout copies are sufficient.' His premise is false. There are 52 people in attendance — Ted double—counted one person who changed seats during the count.

At first glance, Defeasibility seems to give the wrong result about *Handouts* (at least to the extent that we feel inclined to attribute knowledge to Ted): the truth 'It is not the case that there are 53 people in attendance' seems to be just the kind of truth (4) says is a defeater and, hence, a truth that lowers one's justification below the threshold required for knowledge. The problem is that this verdict dispensed by Defeasibility clashes with people's intuition that Ted knows.

Peter has replied to cases like *Handouts* in the following way (Klein, 2017): in those cases, the truth that seems to defeat the subject's justification is itself defeated by something else the subject knows and that is entailed by the alleged defeater. So, Ted knows that he has enough handouts because the epistemic damage 'It is not the case that there are 53 people in attendance' could do to his justification is preempted[20] by his knowledge that there are approximately 52 people in attendance. I think this reply to *Handouts* is promising.[21] The key feature is that something other than the falsehood is what is turning the conclusion in Ted's inference into knowledge.[22]

---

[20] In (Klein, 2017, p.54) Peter says that this latter truth 'restores' Ted's justification for believing that he has enough handouts. I find that confusing — if 'It is not the case that there are 53 people in attendance' is not really a defeater, then there is no damage to be *restored* by this truth. Instead, what seems to be going on here is that the destructive effect of 'It is not the case that there are 53 people in attendance' is *preempted* by Ted's knowledge that there are 52 people in attendance.

[21] I have given a similar reply to *Handouts* myself in (Borges, 2017). See (de Almeida, 2017) for criticism of Peter's way of handling cases of useful falsehoods. For alternative ways of dealing with *Handouts* as well as with other cases of knowledge from non—knowledge, see, among others, (Ball & BlomeTillmann, 2014), (Montminy, 2014), (Luzzi, 2014), and (Schnee, 2015).

[22] It may be suggested that Peter should not have been so quick to assume that Ted is 'epistemically close' to what is obviously entailed by something he knows. After all, closure—deniers such as Fred Dretske and Robert Nozick can reasonably reject the claim about epistemic proximity Peter relies on. This is an important worry about Peter's strategy, but I think it can be mitigated. I do not have space to give it a full treatment here; so, a couple of quick

There might be problems for Defeasibility when we compare its response to *Handouts* with its response to *Marbles/Marbles**. Why can't we say that something Mary knows prevents her knowledge that there is a red and black marble from being (epistemically) defeated, like we did with Ted's knowledge that his one hundred handouts are enough? After all, is it not the case that Mary *saw* that she put a black and a red marble in the bag? Why can't we take this knowledge to preempt the defeating effect of the counterevidence coming from the drawings? Saying that Defeasibility can treat those cases differently because the subject knows the target proposition in one case but not in the other is utterly unsatisfactory. Supposedly, the *analysans* in a correct analysis of the concept of knowledge is more informative than its *analysandum*. But, if in some cases we need first to judge whether the *analysandum* is instantiated before we can learn whether the *analysans* is instantiated, then the analysis under consideration is failing to display that core feature of correct analyses of the concept of knowledge.

This issue generalizes. For example, supposing we want Defeasibility to accommodate the intuitive judgments that the subjects in *Headache* and *Weather* know, while the subject in *Marbles/Marbles** does not know, then we need an account of why the counterevidence in the latter case defeats the subject's justification while the counterevidence in the first two cases does not. Given that in all

three cases the counterevidence is encapsulated in a truth that could, at least in principle, be taken to defeat the subject's justification, it seems arbitrary for Peter to say that justification gets defeated in one case but not in the other. It seems that, in general, there is nothing in Defeasibility itself that allows us to state, in a principled way, when counterevidence epistemically defeats knowledge and when it does not. Call this *the indeterminacy problem* for Defeasibility.

Peter has, of course, addressed the indeterminacy problem in a number of his writings, but I am not convinced his main strategy is fully satisfactory. Let me reconstruct this strategy here.[23] When considering some particularly tricky cases where the subject has a justified true belief but it is not so clear whether there is a defeater preventing her from knowing, Peter suggests that it is a virtue of Defeasibility that it neither necessitates nor precludes the existence of a defeater.[24] To illustrate this point consider the following case:[25]

> *Newspaper*
> A political leader is assassinated and the local newspaper reports the event accurately. The whole island, including Smith, reads about it. He reads the newspaper and believes, on that basis, that the leader has been assassinated. However, the leader's party, fearing a coup, later convinces an otherwise reliable TV channel to televise the false story that a member of the security team, not the leader, has been killed. Although everyone on the island believes the televised piece of fake news, dismissing the 'rumor' that the leader was assassinated, Smith is unaware of the broadcast.

---

[23] Peter used to call this problem the 'problem of misleading defeaters' but he has recently dropped this description of the problem, and for a good reason: since, according to Peter, misleading evidence against what one knows is just counterevidence that *does not* defeat knowledge; calling a misleading defeater a 'defeater' is just confusing, for defeaters, by definition, do *defeat* and the so—called 'misleading defeaters' do not defeat. See (Klein, 2017).

[24] (See, for example, Klein's discussion of the *Mr. Magic* case in (Klein, 2003).

[25] This is a version of a case presented by (Harman, 1973, p.143—144).

Does Smith know that the political leader was assassinated? Peter argues that it is a virtue of Defeasibility that it is compatible with a positive *and* a negative answer to this question. According to Defeasibility, which answers we give depends on whether we take the true proposition *a reliable TV channel said that a member of the security team, not the leader, has been killed* to be a defeater of Smith's justification. If we think Smith knows, then we should *not* consider it to be a defeater; if we do not, then we should not. As I said before, this kind of reply on behalf of Defeasibility is less informative than one would have hoped for. We expect that *analysans* in a correct analysis of knowledge to be more informative than its *analysandum* and that this means, in turn, that finding out whether the latter is satisfied in a case should take explanatory priority over whether the former is satisfied in that case, not the other way around.[26] But this expectation is frustrated by Peter's approach to the indeterminacy problem.[27]

I do not take this problem to be a knockdown argument against Defeasibil-

---

[26]Objection: It is *vague* whether Smith knows or not in this case, and Defeasibility does exactly what we want it to do — it explains *clear cases* of knowledge/ignorance. Reply: The relevant accusation (what I mean by 'less informative than one would have hoped for') is that we look for defeaters only after we decide whether there's knowledge in a relevant case. This gets things backwards, explanatorily speaking. As in science, we want philosophical theories that predict what type of evidence we are likely to find (whether we will find knowledge in a case or not) and not theories that only account for evidence we already have (whether cases that we already know have knowledge/ignorance conform to the theory). My claim is that Defeasibility has poor predictive power. Consider. The claim 'All emeralds are green' tells me that, for any x, if x is not green, then x is not an emerald. This conditional exploits the *analysans* (so to speak) of 'emerald' and suggests a method for finding emeralds (i.e, looking at the color of things and disregarding things that are not green). Similarly, 'All knowledge is undefeated belief' suggests that, for any x, if x is defeated (if there is a defeater), then x is not knowledge. This conditional exploits the *analysans* of 'knowledge' and suggests a method for finding knowledge (i.e., looking for defeaters and disregarding beliefs for which there are defeaters). However, while I can look for non—green things even if I do not know whether there are emeralds present or not, I do not see how I can look for defeaters without first knowing whether there is knowledge present or not. This is not good. It might not 'refute' Defeasibility, but it's a problem that should be addressed, I think. Also, it is not so clear that it is vague whether the subject knows or not in *Newspaper*. Nozick (1981, p.177), for example, thought it was a virtue of his proposed definition of knowledge that it explained 'why we are reluctant to say [the subject in *Newspaper*] knows the truth.' According to him, condition 4 in his definition of knowledge (i.e., if p, then S would believe that p) predicts the right result in that case — that the subject fails to know. What this tells me is that Nozick did not think the subject knew, that he thought that his theory explained why this was the case and that this conformed with what he took to be people's general inclination about the case. Harman (1973, p.145) also argued that Smith did not know and that this was a problem for his account. Thanks to Joao Fett for discussion here.

[27]A similar point applies to some versions of the Grabit case, which was originally presented in (Lehrer & Thomas Paxson, 1969). See (Klein, 1981) for Peter's treatment of the Grabit case.

ity. Rather, I wanted simply to illustrate the kind of challenge any weak view of knowledge faces.[28]

V

I now turn to a recent version of the strong view of knowledge. I will suggest that it too faces important limitations. I will then briefly suggest one way in which these limitations can be remedied.

Maria Lasonen-Aarnio (2010) has recently defended a version of the strong view of knowledge. As such, the view states that counterevidence can never epistemically defeat knowledge. She does, however, want to explain why some have the defeat intuition (i.e., the inclination to think that subjects in cases such as *Marbles/Marbles** have their knowledge epistemically defeated by counterevidence). According to Lasonen-Aarnio, although 'these intuitions are mistaken, subjects who retain knowledge in defeat cases are genuinely criticizable.' In other words, although Lasonen—Aarnio does not think that knowledge can be epistemically defeated, she wants to give an account of why subjects who retain knowledge when confronted with counterevidence deserve to be criticized.

According to her, those subjects are criticizable because they do not follow belief—forming policies that are rational for someone with the goal of acquiring knowledge to adopt. Quite the opposite, says Lasonen—Aarnio; to the extent that those subjects manifest the disposition to retain a belief in the face of counterevidence, they are following an epistemic policy that often prevents false belief

---

[28]In fact, one might even say that this way of looking at the indeterminacy problem makes Defeasibility circular, for whether condition (4) is satisfied depends in part on whether the subject knows. In my view, this limitation of belief—first versions of the weak view of knowledge gives credence to a knowledge first version of the weak view: instead of explaining knowledge in terms of defeat (i.e., what explains why one knows, when one does, is the fact that one's justification for believing is undefeated), we should explain defeat in terms of knowledge (i.e., what explains why one's justification is undefeated is the fact that one knows). I only mention this here, however, as I do not have the space to elaborate on this approach to defeat. See (Baker-Hytch & Benton, 2015) for a similar criticism of *reliabilist* versions of weak views.

from being extirpated by new evidence, for this subject is also disposed to retain false beliefs in the face of new evidence. Hence, for Lasonen—Aarnio, the mistaken intuition that there is *epistemic* defeat in cases such as *Marbles/Marbles** is explained by the fact that we have a tendency to think that when a subject is not acting reasonably, she lacks knowledge (Lasonen-Aarnio, 2010, p.15). That is, we mistakenly take a subject's unreasonability in believing to mean that this subject fails to know what she believes. Lasonen-Aarnio (2010, p.16) says that asking whether a belief is reasonable is a 'useful heuristic' for determining whether this belief amounts to knowledge, but she rejects the notion that one's belief's being reasonable is a necessary condition on knowledge.

Even though this view is initially attractive, its account of the defeat intuition is in tension with central features of any strong view of knowledge. Consider the claim that being disposed to retain knowledge in the face of counterevidence is an indication that one is also disposed to retain belief in situations where one is confronted with counterevidence but one has a false belief. Why think that this is the case? That is, why should we think, with Lasonen—Aarnio, that someone who is disposed to retain knowledge in a circumstance in which counterevidence is present is *ipso facto* disposed to retain a false belief in similar circumstances? For one thing, it is logically possible for a reasonable person to have one disposition without having the other (at least for some cases or classes of cases). For instance, I may be disposed to retain knowledge of simple arithmetic truths in the face of counterevidence while, at the same time, being disposed *not* to retain false beliefs about the same subject matter, for the mistakes are easy to spot with the help of new evidence provided, for example, by a calculator.

Perhaps what is behind Lasonen—Aarnio's suggestion is something like the following line of reasoning. Justification is *fallible*; that is, it is possible for S and S* to be equally justified in believing that p even though S knows that p and S* has a justified but false belief that p. Since justification is fallible and one's

disposition to retain knowledge *and* belief is to a good extent determined by how justified one is, a disposition to retain knowledge indicates a disposition to retain justified false belief, for, given the fallibility of justification, there need not be any difference between how justified one is when one knows that p and when one has a justified false belief that p.

The problem with this argument appears when we try to conjoin the claim that justification is fallible with the claim that knowledge cannot suffer epistemic defeat. Lasonen—Aarnio's view can have one or the other claim, but not both. If knowledge is what the strong view of knowledge says it is, then it cannot be the case that S is as justified as S* in believing that p but S knows and S* falsely believes that p. This entailment holds for the simple reason that, according to any strong view of knowledge, the justification of someone who believes falsely that p may be epistemically defeated by new evidence but the justification of someone who knows that p cannot. But, if S and S* were equally justified in believing p, this difference in how resilient to defeat their justification is should not exist. But it does. Hence, if knowledge is what the strong view of knowledge says it is, then justification is not fallible. The upshot is that Lasonen—Aarnio's account of the defeat intuition loses considerable support. Her account of this intuition relies on the claim that justification is fallible, a claim that does not jive well with strong views of knowledge.

VI

I now suggest a way to provide strong views of knowledge with an explanation of the defeat intuition that does not incur the difficulties I mentioned in the previous section. I too will presuppose that, although quite strong, this intuition is ultimately unsound. My account gives a central role to justification, but it does so in a way that coheres with strong views of knowledge.

Some authors in the knowledge—first literature[29] (and elsewhere) have pointed out that 'justification' most likely does not express one single concept. From our argument in the previous section it follows that a strong view of knowledge has to say that the false beliefs about hands formed by handles BIVs are *not* as justified as our knowledge of hands. What this shows is that if we want to hold a plausible strong view of knowledge and, at the same time, say that the BIV's hand—belief is justified, then 'justified' must mean one thing when applied to the BIV's hand—belief and another when applied to our knowledge of hands.

One way to capture the relevant different senses of 'justified' in play here is the following. First, we may say that the sense in which I am justified is the same sense in which I know: I see that these are my hands, and this entails the truth of my belief that I have hands. In fact, it is hard to think of a better reason to believe one has hands than having seen that some things are one's hands![30] Secondly, we may say that the sense in which a BIV is justified in believing she has hands is the sense in which she has an excuse to believe she has hands, and that she should therefore not be blamed for doing so — given how things seem to her, she should be highly confident that she has hands. In fact, it is hard to think of a better excuse to believe one has hands (when one has no hands) than its seeming to one as if one has hands![31]

According to this picture, if we are talking about justification as the thing that is epistemically distinctive about knowledge, then we are talking about 'being justified' as having what Fred Dretske (1971) called a 'conclusive reason,' a reason that would not be the case unless the target belief were also the case. I would not have seen that I have hands unless I had hands. Justification, in the sense in which we say that the BIV is justified in believing she has hands, is an

---

[29] For example, (Sutton, 2007), (Williamson, 2011), (Littlejohn, 2012), among others.

[30] This is similar to some versions of *epistemological disjunctivism*. For one such view, see (Pritchard, 2016).

[31] Williamson (2015) and Littlejohn (2012) make a similar distinction.

entirely different beast, however. For one, it is not the case that its seeming to her that she has hands is a conclusive reason for her to believe she has hands (i.e., it is not the case that things would not seem to her the way they do unless she had hands). If we use this distinction (i.e., 'justified' as knowing vs. 'justified' as having an excuse) in an explanation of why some of us have the defeat intuition, we get the following simple story: in cases where knowledge is allegedly defeated by counterevidence, it seems to the subject as if what she knows is false; this change in how things seem to her, in turn, leads her to blamelessly form a belief in a false proposition (i.e., the negation of what she knows) and to stop believing what she knows.

There are two crucial differences between those two senses of 'justified.'

Firstly, having a conclusive reason to believe is a necessary condition on knowledge, but having an excuse isn't — I do not need to have an excuse to believe that p if I have a conclusive reason to believe that p. It is true, however, that there is something wrong, epistemically speaking, with someone who retains her knowledge that p when she has a strong enough excuse to believe that not—p — cases like *Marble/Marble** show this much. Mary knows at first that there is a red and a black marble in the bag. It is because of the dynamic evolution in how things seem to Mary that she is (at first) justified in believing that there is a red and black marble in the bag — in the 'having an excuse' sense of that term — and, then, as she draws a red marble one thousand times she stops being justified.

Secondly, as *Marbles/Marbles** makes clear, justification, understood as having an excuse, can be defeated by new evidence, but justification, understood as having a conclusive reason, cannot. As Mary draws a red marble one thousand times, it stops seeming to her as if there is a red and a black marble in the bag. This is not what happens to Mary's reason: the reason she has for believing that there is a red and a black marble in the bag is the same throughout the case (i.e., she saw herself putting those marbles in the bag). The changes in how things seem

to the subject is what explains why counterevidence sometimes defeats knowledge: when counterevidence to what one knows is strong enough, it makes the known proposition seem false to one, and this, in turn, causes one to decrease one's confidence (and to lose one's belief) in the known proposition; and, at the same time, causes one to increase one's confidence (and to form a belief) in the denial of the known proposition.

As promised, this version of the strong view of knowledge takes counterevidence to defeat knowledge only via the loss of belief, not via the loss of one's justification to believe. In other words, there is only psychological defeat. The defeat intuition is an illusion: we take Mary's knowledge to be epistemically defeated because we recognize that, *from Mary's perspective*, it seems as if there is only a red marble in the bag. In general, the degree to which one is inclined to think that subjects in situations such as Mary's have their knowledge epistemically defeated by counterevidence roughly matches how excusable one thinks it is for subjects in those situations to 'change their mind' about the relevant proposition.[32]

References

Baker-Hytch, M., & Benton, M. A. (2015). Defeatism Defeated. *Philosophical Perspectives*, *29*, 40-66.

Ball, B., & BloméTillmann, M. (2014). Counter closure and knowledge despite falsehood. *Philosophical Quarterly*, *64*, 552-568.

Borges, R. (2015). On Synchronic Dogmatism. *Synthese*, *192*(11), 3677-3693.

Borges, R. (2017). Inferential Knowledge and the Gettier Conjecture. In R. Borges, C. de Almeida, & P. D. Klein (Eds.), *Explaining Knowledge: New Essays on the Gettier Problem* (p. 273-291). Oxford University Press.

de Almeida, C. (2017). Knowledge, Benign Falsehoods, and the Gettier Problem. In R. Borges, C. de Almeida, & P. D. Klein (Eds.), *Explaining Knowledge: New Essays on the Gettier Problem* (p. 292-311). Oxford University Press.

Dretske, F. (1971, May). Conclusive Reasons. *Australasian Journal of Philosophy*, *49*(1), 1-22.

Dretske, F. (1981). The Pragmatic Dimension of Knowledge. *Philosophical Studies*, *40*(3), 363-378.

Fine, G. (2004). Knowledge and True Belief in the Meno. In D. Sedley (Ed.), *Oxford studies in ancient philosophy* (Vol. XXVII, p. 41-81). Oxford University Press.

Greenough, P., & Pritchard, D. (Eds.). (2009). *Williamson on knowledge*. Oxford University Press.

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

Hintikka, J. (1962). *Knowledge and belief*. Cornell University Press.

Klein, P. D. (1971). A Proposed Definition of Propositional Knowledge. *The Journal of Philosophy*, *68*(16), 471-482.

Klein, P. D. (1981). *Certainty: A Refutation of Scepticism*. University of Minnesota Press.

Klein, P. D. (2003). Knowledge is True, Non-Defeated Justified Belief. In S. Luper (Ed.), *Essential Knowledge.* Longman.

Klein, P. D. (2008). Useful False Beliefs. In Q. Smith (Ed.), *New essays in epistemology* (p. 25-63). Oxford University Press.

Klein, P. D. (2017). The Nature of Knowledge. In R. Borges, C. de Almeida, & P. D. Klein (Eds.), *Explaining Knowledge: New Essays on the Gettier Problem* (p. 35-56). Oxford University Press.

Kripke, S. (2011). *Philosophical Troubles.* Oxford University Press.

Lasonen-Aarnio, M. (2010). Unreasonable Knowleedge. *Philosophical Perspectives, 24,* 1-21.

Lehrer, K., & Thomas Paxson, J. (1969, April). Knowledge: Undefeated Justified True Belief. *The Journal of Philosophy, 66*(8), 225-237.

Littlejohn, C. (2012). *Justification and the Truth-Connection.* Cambridge University Press.

Luzzi, F. (2014). What Does Knowledge-Yielding Deduction Require of Its Premises? *Episteme, 11*(3), 1-15.

Malcolm, N. (1952). Knowledge and Belief. *Mind, 61*(242), 178 - 189.

Montminy, M. (2014). Knowledge despite falsehood. *Pacific Philosophical Quarterly, 44*(3-4), 463-475.

Nozick, R. (1981). *Philosophical explanations.* Belknap Press of Harvard University Press.

Pritchard, D. (2016). *Epistemic Angst: Radical Skepticism and the Groundlessness of Our Believing.* Princeton University Press.

Schnee, I. (2015). There is no knowledge from falsehood. *Episteme, 12*(1), 53-74.

Sorensen, R. (2012). Epistemic Paradoxes. In *Stanford encyclopedia of philosophy.* Stanford University. Retrieved January 2013, from `http://plato.stanford.edu/archives/spr2012/entries/epistemic-paradoxes/`

Sutton, J. (2007). *Without Justification*. MIT Press.

Warfield, T. A. (2005). Knowledge from Falsehood. *Philosophical Perspectives*, *19*(1), 405-416.

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

Williamson, T. (2011). Knowledge First Epistemology. In S. Bernecker & D. Pritchard (Eds.), *The routledge companion to epistemology* (p. 208-218). Routledge.

Williamson, T. (2015). Justifications, Excuses, and Sceptical Scenarios. In J. Dutant & D. Dohrn (Eds.), *The new evil demon*. Oxford University Press.