Philos Stud (2022) 179:93–108
https://doi.org/10.1007/s11098-021-01652-x

The diachronic threshold problem

Rodrigo Borges[1]

Accepted: 27 April 2021 / Published online: 31 May 2021

Abstract The paper introduces a new problem for fallibilist and infallibilist epistemologies—the diachronic threshold problem. As the name suggests, this is a problem similar to the well-known threshold problem for fallibilism. The new problem affects both fallibilism and infallibilism, however. The paper argues that anyone who worries about the well known problem for fallibilism should also worry about this new, diachronic version of the problem.

Keywords Threshold problem · Knowledge · Justification

1 Introduction

Fallibilism is often criticized for lacking a non-arbitrary way of explaining the degree of justification that is required for knowledge. This is known as *the threshold problem for fallibilism*.[1] This problem for fallibilism is synchronic in nature, since it deals with the arbitrariness of the threshold one must satisfy *at a particular time* in order to count as having knowledge-level justification. This paper argues that those who take this problem seriously should also take seriously a diachronic version of the problem that concerns the arbitrariness of the threshold one must satisfy during a period of time in order to count as having knowledge-level justification in that period of time. As it will become clear below, the diachronic version of the problem grows out of the same fundamental intuition behind its sychronic twin: the intuition that arbitrarily small changes in how much evidence one has for p do not make a

[1] See, for example, Dretske (1981), Hetherington (2006), Brown (2014) and Hannon (2017).

Rodrigo Borges
epistemen@icloud.com

[1] University of Florida, Gainesville, USA

difference to the question of whether one knows that p or not. As we will see, although fallibilists and their detractors (i.e., infallibilists) take the synchronic problem seriously, they do not yet devote any attention to the related diachronic problem. This is unfortunate, since it can be shown that the diachronic version of the problem threatens not only fallibilist views, but also some of the most promising infallibilist view in the market. Given this, the main goal of the paper is to establish the following conditional claim: *if* the synchronic threshold problem is a serious problem, then so is the diachronic threshold problem. I will assume that readers believe that the the well-known version of the threshold problem poses a serious problem for fallibilism. Hence, this paper offers little to support the existence of that problem. My efforts will be focused on explaining why someone who cares about that well-known problem already has sufficient reason to care about the diachronic problem as well. If I am right about this, then fallibilists and infallibilists alike have a new problem with which to contend. Not only that, since my argument shows that infallibilism suffers from the diachronic threshold problem, the claim that infallibilism evades the threshold problem is not entirely true. This is an important result, for infallibilists claim that evading the threshold problem is one of the main reasons anyone has for accepting their view.

Lastly, a note of fair warning. This is a paper about a new epistemological problem and the epistemological theories it afflicts. This is not a paper about how to solve this problem, or a survey of the ways in which the existent epistemological theories might try to solve this problem. Some responses are considered in Sect. 4, but they are not the focus of the discussion. The diachronic threshold problem, and the epistemological theories it afflicts—that is the focus of this paper.

I will proceed in the following way. Section 2 handles conceptual preliminaries. Section 3 discusses the traditional, synchronic, threshold problem for fallibilism. A particular case is used to illustrate this problem, and the issue behind the threshold problem is made explicit. The section also briefly discusses a couple of things fallibilists have said in reply to this issue. Section 4 discusses the claim, made by some infallibilists, that their view is preferable to fallibilism because it solves the threshold problem. It is argued that this conclusion is premature since infallibilism is afflicted by the diachronic version of the problem. Section 5, the conclusion, draws some general lessons from our discussion.

2 Preliminaries

In what follows I will represent degrees of justification numerically (e.g., ‘S’s belief that p is .95 justified’). I don’*t* mean this to be a deep point about how we should understand justification. Representing degrees of justification numerically simply makes the discussion I want to have significantly simpler.[2] I trust it will be clear that, if one wanted to, the thrust of what I want to say about justification and the

[2] See, among others, Williamson (2000), Shogenji (2012) and Kotzen (2019) for other philosophers who represent justification numerically in more or less the same way I am representing it here. See Achinstein (2003) for a dissenting view.

threshold problem could be put into purely qualitative terms (e.g., 'S is partially/fully justified in believing that *p*').[3] Also, and unless otherwise noted, I use 'justified' (in contrast with 'partially justified', etc.) to refer to the degree of justification (whatever that may turn out to be) that is, *other things being equal*, sufficient for knowledge. In other words, my discussion of knowledge and of knowledge-level justification will ignore the Gettier Problem, and any condition one might wish to impose on knowledge in order to avoid the Gettier Problem.[4]

One benefit of representing degrees of justification numerically is that it helps us make precise claims about justifying factors (e.g., 'The fact that *a* looks red justifies *S* in believing that *a* is red'). Those claims may be understood as claims about the degree to which this or that factor raises the probability that a proposition, *p*, is true. In other words, if *e* is a justifying factor for S's belief that *p*, then $P(p \mid e) > P(p)$.[5] This way of representing justification also allows us to represent justifying factors that contribute negatively to justification, such as defeaters (e.g., 'The fact that there is a red light shining on the objects around S defeats her justification for believing that *a* is red'). These claims may be understood as claims about the degree to which this or that factor lowers the probability that *p* is true. In other words, if d is a defeating factor for S's belief that *p*, then $P(p) > P(p \mid \mathrm{d})$.

This way of representing positive and negative justifying factors helps us underscore the gradable character of justification and defeat. One's belief that *p* may be made more or less (or only *slightly* more or *slightly* less) justified by some factor, f. It is because justification comes in degrees that it makes sense for us to say things like 'Liz has some justification for believing that the butler did it, but she is not fully justified in believing that.' If this sentence is appropriately used to describe Liz's situation, this is because justification is gradable, and Liz might have *some* (but not *all*) of what is needed for a justified belief (i.e., a belief that, if true, amounts to knowledge, other things being equal). It implies that Liz has some reason or evidence to believe that the butler did it, but that she is not (yet) entitled to do so. Maybe she is a detective, and she has only just begun her murder investigation. It also makes sense for people to say things like 'Liz seems justified in believing that the butler did it, but the scuff marks in the kitchen floor suggest otherwise.' This sentence may, for example, appropriately describe Liz's epistemic state at the end of her investigation, when she is in possession of evidence that establishes the butler's culpability beyond a reasonable doubt. The scuff marks provide Liz with *some* (but not *all*) of what is needed for the defeat of her justification.

[3] See Hetherington (2006) for an apt discussion of the synchronic threshold problem in qualitative terms.

[4] Most epistemologists think that a correct definition of knowledge must be Gettier-proof. I discuss the Gettier Problem in Borges (2017) and Borges (2020a).

[5] The 'P( )' function refers here to what is sometimes called *evidential probability* (broadly speaking): when we say that *p* is probable given *e*, for example, we are saying something like 'Any reasonable person who considers *p* in light of *e* carefully enough will find believing *p* to be more reasonable than believing not-*p*.' 'Reasonable' and 'more reasonable than' are primitive epistemological concepts as in Chisholm (1966).

These are common epistemic phenomena that emerge because of the gradable nature of justification. They also provide our numerical representation of justification with an intuitive rationale.[6]

3 Fallibilism and the threshold

The central tenet of fallibilism, *qua* fallibilism, is that it is possible for S to know that p at time t even if S's justification for believing that p does not guarantee the truth of p at t.[7] In symbols (where '$\Diamond$' stands for 'it is conceptually/metaphysically possible that,' '$\Box$' stands for 'it is conceptually/metaphysically necessary that,' 'Kp' abbreviates 'S knows that p,' and 'e' stands for one's total evidence):

1 $\Diamond$ [Kp $\wedge$ P(p | e) < 1]

Moreover, the fallibilist must assume some lower bound for the type of justification required for knowledge that is higher that .5. This follows from the fact that P($p \vee \neg p$) = 1, according to the axiom of normalization and the intuition that one does not know that p if $\neg p$ is as probable as p on one's total evidence. In symbols:

2 $\neg\Diamond$ [Kp $\wedge$ P(p | e) $\leq$ .5]

Because the fallibilist accepts 1 and 2, she also accepts the claim that there is some threshold for justification, n, such that any true belief that is justified to degree n amounts to knowledge (other things being equal).[8] In other words, 3 follows from 1 and 2:

3 $\Box$ [Kp $\supset$ P(p | e) = n] (where .5 < n $\leq$ 1)

Of course, once the fallibilist is settled with 3, the question becomes: which value should n take? Suppose the fallibilist says that the value of n should be at least .99 ($n \geqslant .99$). Given this account of the threshold for justification, one may construct a case to show that this is a seemingly arbitrary choice. Suppose Liz puts 98 black marbles and 2 red marbles in a bag. Suppose also that she knows that this is case. Further, suppose that she also knows that the bag has no internal pockets or any other abnormality. Liz shakes the bag, places her hand inside and grabs one marble. As she does that, she comes to believe (truly), at t, that she is holding a black marble. Now, most people would say that Liz is justified in believing that she is holding a black marble. After all, she knows that a disproportionate amount of marbles are black, and she knows that she is holding a normal bag. The problem is that the fallibilist view under consideration has to reject this intuition as misleading: the probability that Liz is holding a black marble, conditional on her evidence, is

[6] See Kotzen (2019) for a similar account of justification and defeat.

[7] See Dougherty (2011) for a discussion of fallibilism that goes beyond my focus on fallibilism *qua* fallibilism.

[8] Remember: we are ignoring the Gettier Problem. We are also ignoring global forms of skepticism (e.g., Cartesian skepticism).

'only' .98, and the fallibilist threshold justification, *n*, is set at .99. In other words, Liz does not satisfy the threshold for justification suggested by the fallibilist.

This is quite an unfortunate result for the fallibilist, since it seems to make the threshold of knowledge-level justification quite arbitrary. It seems problematic for the fallibilist to say that Liz is not knowledge-level justified, but that she would have been knowledge-level justified if one of the red marbles had been black! Intuitively, such a *small* change in her evidence should not make this much of a difference to her knowledge. This is akin to saying that a person can go from being clearly not-bald to being clearly bald by losing a single hair.[9]

Now, the fallibilist will not be able to dismiss the charge of arbitrariness by assigning a different value to *n*. This is because the result from the previous paragraph generalizes: for any plausible value *n* might take, there is a marble-like case that suggests that *n* is arbitrary. For any plausible value *n* might take there is a case like the one in the previous paragraph and in which there are *n*-1 black marbles and 101-*n* red marbles in Liz's bag. This generalization of Liz's case is what goes by the name of the *threshold problem* for fallibilism in the epistemological literature.[10] Here is Fred Dretske discussing this issue (Dretske 1981, pp. 363–364):

> Philosophers who view knowledge as some form of justified true belief are generally reluctant to talk about this implied threshold of justification. Just how much evidence or justification ... is enough to qualify as an adequate, a full, or a complete justification [(i.e., the sort of justification required for knowing)]? ... any threshold less than 1 seems arbitrary. Why, for example, should a justification of 0.95 be good enough to know something when a justification of 0.94 is not adequate?

Stephen Hetherington (2006, p. 42) also underscores the challenge fallibilists face; for him, the problem is compounded by the fact that most epistemologists *are* fallibilists:

> ... epistemologists in general face a conceptual challenge of either removing or disarming that vagueness in any fallibilist conception of knowledge. Most epistemologists need to show why that vagueness does not undermine all putative fallibilist theories of knowledge.

These quotes illustrate the current understanding of the threshold problem. In particular, they illustrate the fact that this problem is *structurally similar* (but not identical) to the problem of vagueness.[11] The threshold problem for fallibilism may be understood as the problem of providing a non-arbitrary account of what fixes the

[9] But see a few paragraphs down (and footnote 11) for more on the similarity between the vagueness of 'know' and 'bald.'

[10] See, among others, Dretske (1981), Hetherington (2006), Bonjour (2010), Brown (2014) and Hannon (2017).

[11] The claim that the problems of vagueness and of the threshold are the same is controversial. The idea that those problems are structurally similar is more widely accepted in the literature on the threshold problem. See Hannon (2017, footnote 4) for a similar view on the relationship between the two problems.

threshold of non-conclusive, knowledge-level justification in a way that is consistent with our intuitions about the nature and value of (fallible) knowledge.[12]

But in spite of the attention devoted to this problem for fallibilism, the literature has so far neglected the fact that fallibilism suffers from a related but distinguishable problem. In order to see that, we need only remind ourselves that most fallibilists believe that justification is not only fallible (i.e., a belief may be justified, yet fail to be true), but that it is also *defeasible* (i.e., a belief may be justified now, yet fail to be justified later). That is, fallibilists allow for *future* new evidence to lower one's *current* degree of justification below the threshold, *n*, required for knowledge.[13] Gilbert Harman describes this form of defeasibilist fallibilism in the following passage (Harman 1973, p. 149):

> Since I now know that [*p*], I now know that any evidence that appears to indicate something else is misleading. That does not warrant me in disregarding any further evidence, since getting that further evidence can change what I know. In particular, after I get such further evidence I may no longer know that it is misleading. For having the new evidence can make it true that I no longer know that new evidence is misleading.

Although Harman discusses a particular item of knowledge (i.e., knowledge that evidence against *p* is misleading), and how it may be defeated by future new evidence, the point is fully general: *any* item of knowledge may be defeated by future new evidence suggesting that what one knows is false.[14] Even if I know that my car is in the parking lot, this knowledge may be defeated, say, by the sincere (but mistaken) testimony of the sheriff who tells me that one of his deputies found my car in the bottom of a lake.

Thus, according to most fallibilist views, one may be justified in believing that *p* at a time *t*, but cease to be justified at a later time, *t**. This means that at some point between *t* and *t**, the degree of justification in support of one's belief that *p* fell below the threshold for justification, *n*, required for knowledge. To see that, consider the following version of the marble case I mentioned a few paragraphs ago. Suppose Liz has a bag with one red marble and one black marble. She knows that. And she knows that the bag has no abnormalities (i.e., no pockets, etc.). Starting at time *t*, Liz makes draws with replacement from this bag. To her surprise, by the time she stops drawing marbles from her bag, at *t**, Liz has drawn a red marble two hundred and fifty times in a row.[15] Since they are also defeasibilists, fallibilists will say that, at *t**, Liz no longer knows that there is a red *and* a black marble in the bag. According to them, each time Liz draws a red marble she acquires some (however

[12] See Hannon (2019, p. 57) for a similar characterization of the problem. Of course, this way of putting the problem should not be taken to imply that the fallibilist can solve the threshold problem by saying that the threshold is 'not precise,' or 'too vague.' I agree with Hannon (2019, p. 64), Bonjour (2010) and others that the fallibilist does not solve the problem in this way.

[13] There are innumerable papers discussing and/or endorsing this view, but prominent and recent instances are de Almeida (2017), Klein (2019) and Kotzen (2019).

[14] See also Hawthorne (2004, p. 73) for the same point.

[15] This case is adapted from Williamson (2000, p. 205).

small) evidence that the bag has two red marbles instead of one red marble and one black marble. After having drawn a red marble an x number of times in the period between *t* and *t**, Liz will have accumulated enough (misleading) evidence to lower her degree of justification below the threshold, *n*, for knowledge-level justification.[16] But what is so special about x consecutive draws of a red marble that it defeats Liz's justification for knowing that there is a red and a black marble in the bag? What about x−1 draws of a red marble? Why is one single draw of a red marble allowed to make such a big difference, thus turning Liz's justified true belief into an *unjustified* (i.e., not knowledge-level justified) true belief? Intuitively, such a *small* difference in her evidence should not make such a *big* epistemic difference. Again, this seems similar to saying that a person can go from being clearly not-bald to being clearly bald by losing a single hair. What is more, the issue applies to any number of draws, x, of a red marble one cares to select.[17] This may be called *the diachronic threshold problem.* This is the problem of providing a non-arbitrary

[16] If you think that knowledge is vulnerable to new evidence in the way suggested by the fallibilist, but you do not think that two hundred and fifty consecutive draws of a red marble are enough to defeat Liz's justification, then change the case so that the number of consecutive draws of a red marble matches your epistemological sensibilities. Alternatively, if you think that there is no number of drawings of a red marble that is sufficient to lower Liz's justification below *n*, then you might be an infallibilist. There is nothing wrong with being an infallibilist, but then this argument is not designed to convince *you* of anything. The impatient infallibilist should simply skip to the next paragraph where I argue that fallibilism and defeasibilism belong together.

[17] To the extent that the marbles case deals with the border between ignorance and knowledge, it resembles some cases discussed in the literature on inductive knowledge. For example, Cian Door, Jeremy Goodman and John Hawthorne discuss a case where the subject seems to learn that a certain double-headed coin is not fair by flipping it repeatedly and seeing it land heads each time. According to Door et al., '[i]n any such case, there must be a first flip of the coin after which you are in a position to know that the coin is not fair' Door et al. (2014, p. 283). Similarly, Bacon (2020) asks how one could come to know, after observing emeralds e_1 ... e_n to be green, that a law explains all of one's observations, rather than mere chance. Bacon asks how one observation could make a difference. The cases discussed by Door et al. and Bacon are similar to the marbles case, in that they all feature the inductive and gradual crossing of the border between ignorance and knowledge. But, even though the cases are similar in this way, they are being used to ask importantly different questions. On the one hand, Door et al. and Bacon's cases are being used to probe *fallibilism* with the question 'How is inductive knowledge possible in a fallibilist framework (i.e., in a framework in which knowledge-level justification does not require probability 1)?' On the other hand, I use the marbles case to probe *defeasibilism* (in its fallibilist and infallibilist varieties) with the question 'How is defeat possible in a defeasibilist framework (i.e., in a framework where new misleading evidence is allowed to defeat existing knowledge)?' The result is that the cases Door et al. and Bacon discuss are similar to the marbles case in broad outline, but they are being used to raise different questions about theories that are trying to explain different epistemological phenomena. While the theoretical target in the Door et al. and Bacon cases is the crossing of the border between ignorance and knowledge in the *ignorance-to-knowledge* direction, the theoretical target in the marbles case is the crossing of the same border but in the opposite direction; i.e., in the *knowledge-to-ignorance* direction. The direction in which the crossing occurs matters (among other reasons) because one may cross into ignorance (from knowledge) with respect to p by suspending judgment about p, but one cannot cross into knowledge that p (from ignorance about p) in the same way. The use I make of the marbles case poses a new challenge to defeasibilism, a challenge that threatens to undercut the claim that infallibilists need not worry about the problem of the threshold. If am right, some prominent infallibilists need to worry about a version of the problem, a version of the problem that comes into focus as one thinks carefully about Liz's crossing of the threshold between ignorance and knowledge in the *knowledge-to-ignorance* direction. The traditional threshold problem for fallibilism, as well as the cases in Door et al. and Bacon, do no such thing. Many thanks to a reviewer for bringing these cases to my attention.

account of what fixes the threshold of (conclusive or non-conclusive) knowledge-level justification in a way that is consistent with our intuitions about the nature and value of (fallible or infallible) knowledge. This much seems clear: *if* setting the threshold for justification to degree *n* raises the specter of arbitrariness in the way suggested by the traditional, synchronic, threshold problem, *then* setting the threshold for justification to degree *n* also raises the specter of arbitrariness suggested by the diachronic threshold problem.[18]

But, in what sense is this problem diachronic in nature?[19] As state by Harman (see above) and others, defeasibilism is an account of *belief revision*; that is, an account of an inferential, psychological process. In that sense, the diachcronic threshold problem is essentially diachronic to the extent that it is a problem afflicting the view of belief revision fallibilists and (some prominent) infallibilists accept (i.e., defeasibilism). The process of belief revision—which some like to model using (some version of) the principle of contitionalization—has to do with replacing S's probability distribution 'prior' to her finding new evidence with S's probability distribution 'posterior' to her finding this evidence. What is currently overlooked in the literature about the threshold problem is the fact that the defeasibilist view of belief revision gives rise to a version of the threshold problem.

One might complain that the diachronic version of the problem is not elicited by fallibilism alone, but by the combination of fallibilism and defeasibilism. The idea behind this complaint is that the fallibilist may avoid the diachronic version of the problem by jettisoning defeasibilism. The problem with this suggestion is that fallibilism makes little sense without defeasibilism. Defeasibilism is fallibilism *spread over time*.[20] It would make little sense for the fallibilist to say that S is not certain that *p* at *t*, but that as new evidence suggesting that $\neg p$ roles in S remains

[18] One might worry that the case, as described, does not work because of the 'mere probabilistic nature' of the misleading evidence. However, one may change the case so that this worry disappears but the result remains (i.e., Liz's knowledge is defeated by new evidence). Consider. Liz is drawing marbles with replacement from a normal, opaque bag. She takes careful notes describing each draw. She put a red and a black marble in the bag herself, and she knows that the bag has no abnormalities (pockets, etc.). After a few draws with replacement, her trustworthy butler, Paul, walks into the room and tells Liz that she should not drink from the bottles of water he bought the day before, since he just found out that they are tainted with a substance that causes prolonged periods of hallucination. Chief among them is color confusion: things that are not really red, look red. Now, what Paul says is true (the bottles he bought contain the hallucinogen), and Liz did drink the tainted water. However, what neither one of them knows is that Liz is iron-deficient, and that iron-deficient individuals are not affected by the hallucinogen in the water. Now, a skeptic might object that Liz did not lose any knowledge because of Paul's testimony, since she never had any knowledge to begin with—'Liz is in a Gettier case,' he might try to convince us, 'similar to the Fake Barn case.' I do not find this claim plausible, since both Barney and Liz strike me as knowing the target proposition.' However, I know that this will sound controversial to many (but see Sosa (2007) and Gendler and Hawthorne (2005) for a similar view of Barn cases). So, instead, I suggest that the skeptic change the case so as to make Paul's testimony false but justified. In that case his testimony is misleading because what he says is false, while in the original case his testimony is misleading because it suggests something false (namely, that Liz is hallucinating red objects). Either way, Liz knows at first but not after Paul's testimony; but, in this latter version of the case we avoid Barn-like worries. Thanks to John Biro for discussion here.

[19] Many thanks to a reviewer for prompting me to be explicit about this issue.

[20] Of course, we can also understand defeasibilism *synchronically* (i.e., one is justified in believing that p at time, t, only if there is no defeater of one's justification for p at t). That is not the relevant version of

equally as certain that p is true as she was before. What *makes* sense for the fallibilist to say is this: since S is not-certain that p, at t, she must allow new evidence suggesting that $\neg p$, to lower her degree of certainty in p. How could S be following the evidence if she failed to do that?

This point may be put in a slightly different way as well. The idea animating the fallibilist view of the flux of evidence is that our limited and spotty access to the world generates little to no certainties, our evidence almost always leaving un-eliminated possibilities in which the opposite of what we believe is the case. But, if fallibilism is motivated by the idea that little of what we know amounts to certainty, then rejecting defeasibilism amounts to rejecting the spirit, if not the letter, of fallibilism. If my belief that p is not made certain by my evidence *now*, then how could I rationally reject as misleading *future* new evidence suggesting that $\neg p$ is true? I might be able to rationally reject some future evidence as misleading, for some p's, some of the time. But, clearly, the idea that I will be in a position to do so rationally most of the time, given fallibilism, seems ludicrous. The fallibilist, *qua* fallibilist, is not certain of most things, remember?

4 Infallibilism and the threshold

According to infallibilists such as Fred Dretske, Laurence Bonjour, and others, fallibilists are nowhere near solving the traditional, synchronic, threshold problem. According to them this fact about fallibilism gives one a strong reason to join the infallibilist ranks. Here are Bonjour and Dretske, respectively, making this point.

> ...there simply is no well-defined, intellectually significant concept of knowledge fitting [the fallibilist] conception: none that can be genuinely found in common sense or indeed can be constructed or stipulated in a satisfactory way. [The fallibilist account of knowledge] is, I am suggesting, a philosophical myth.[21]

> Examples ... suggest ... that the absolute, non-comparative character of knowledge derives from the absoluteness, or conclusiveness of the justification required to know. If I know that [p, then] ... I must already have an optimal, or conclusive justification [for p] (a justification at the level of 1)...[22]

Thus, the central tenet of infallibilism, *qua* infallibilism, is that, necessarily, if S knows that p at time t, then S's justification for believing that p guarantees the truth of p at t.[23] In symbols:

Footnote 20 continued
defeasibilism being discussed here, however. For synchronic defeasibilism, see, among others, Lehrer and Thomas (1969), Pollock and Cruz (1999) and Klein (2008).

[21] Bonjour (2010, p. 59).

[22] Dretske (1981, p. 364).

[23] This characterization of infallibilism is good enough for my purposes here, since it is a *sine qua non* for any infallibilist view. However, much more needs to be said in order to make infallibilism plausible,

4 $\Box\,[\mathrm{K}p \supset \mathrm{P}(p \mid e) = 1]$.

In the same way that it made sense to ask the fallibilist why she thinks that any particular *n* is the correct value for the justification threshold, it seems to make perfect sense to ask the infallibilist why she thinks that $n=1$ is the correct threshold for justification. Given what Bonjour, Dretske and others have said about the matter, part of the reason why *n* should be 1 is the absence of a non-arbitrary reason to think that *n* should take any value different than 1. 'This is the lesson we learned from the threshold problem,' says the infallibilist.[24]

There is also another, less often discussed tenet of infallibilism: that *no item of knowledge is immune to defeat by new evidence*. 'No matter how certain a proposition is on one's evidence,' the infallibilist will say, 'accumulating evidence suggesting that the proposition is false will eventually reach critical mass and lower one's degree of justification below the threshold required for knowledge '(i.e., 1).' Here is a prominent infallibilist, Timothy Williamson[25] making this very point:[26]

> 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.[27] (Williamson 2000, p. 206)

Because the infallibilist allows for new evidence to push one's justification below the threshold required for knowledge, she accepts defeasibilism.[28] This is a rather

---

Footnote 23 continued
as a whole. For a careful presentation of infallibilism that aims to do just that, see, e.g., Dretske (1971), Williamson (2000), Neta (2009) and Pritchard (2016).

[24] It might be argued that setting the threshold for justification at 1 is as arbitrary as any other value between .5 and 1 (at least at a first glance). Consider: Liz puts 99 black marbles and 1 red marble in a bag. She knows this to be the case; and, she knows that the bag has no internal pockets or any other abnormality. Liz shakes the bag, places her hand inside and grabs one marble. As she does that, she comes to believe (truly) at *t* that she is holding a black marble. Many would say that Liz is justified in believing that she is holding a black marble. However, this verdict is incompatible with infallibilism since, the probability that she is holding a black marble, given her evidence, is .99 instead of the required 1. However, because the infallibilist might reasonably complain that she is being accused of holding her own view, I will not pursue the issue any further.

[25] According to Williamson (2000), if one knows that *p*, then *p* is part of one's evidence. This means that *p* has probability 1 given S's evidence whenever S knows that *p*. So, for Williamson, knowledge-level justification is infallible.

[26] Dretske allows for misleading evidence to undermine knowledge; however, for him, misleading evidence leads to the loss of knowledge because it undermines *belief*, not because it undermines *justification*. According to him, '... if a person really does believe that [q (where q is incompatible with *p*)], aside from the question of whether or not this belief is reasonable, he surely fails to have the kind of belief requisite to knowing [that *p*]. He certainly doesn't think he knows [that *p*]. I do not know exactly how to express the belief condition on knowledge, but it seems to me that anyone who believes (reasonably or not) that he might be wrong fails to meet it' (Dretske 1981, p. 376). In the main body of the paper the focus is on infallibilist views, such as Williamsonian's, that allow for the possibility of defeat of justification.

[27] Williamson uses 'evidence' where I used 'knowledge.' The change is justified because the context in which this passage appears is one in which Williamson is defending the view that evidence *is* knowledge.

[28] Exceptions are Malcolm (1952), Hintikka (1962), and, of course, Descartes (2008).

surprising result, since, as I argued above, defeasibilism is nothing but fallibilism spread over time. This is surprising also because it implies that defeasibilim is more widespread than fallibilism, since the set of defeasibilist views includes fallibilist views *and* infallibilist views. Perhaps even more surprising is the fact that this feature of infallibilist views (i.e., their defeasibilism) has received little to no attention in the epistemological literature. This is a grave oversight, for the coupling of infallibilism and defeasibilism leads straight to a dilemma whose horns are, on the one hand, dogmatism and, on the other hand, the diachronic threshold problem.[29]

We may illustrate how this dilemma arises for the infallibilist by revisiting Liz and her marbles. Liz has a bag with one red marble and one black marble. She knows that. And she knows that the bag has no abnormalities (i.e., no pockets, etc.). Starting at time *t*, Liz makes draws with replacement from this bag. To her surprise, by the time she stops drawing marbles from her bag, at *t**, she has drawn a red marble two hundred and fifty times in a row. What should the infallibilist say about this case? There are only two live options here:

a. Liz knows at *t*, and at *t**, that there is a black and a red marble in the bag.
b. Liz knows at *t*, but not at *t**, that there is a black and a red marble in the bag.

The infallibilist who accepts (a) might say that Liz knows that, although statistically unlikely, a sequence of two hundred and fifty drawings of a red marble is possible, and that those drawings provide her with no reason to doubt her knowledge that the bag has a red and a black marble. She might say 'Well, *if*, in light of the consecutive drawings of a red marble, Liz comes to believe that she might have made a mistake, *then* she is no longer justified (and, hence, no longer knows) that there is a red and a black marble in the bag; otherwise, she continues to know.' In other words, if Liz holds fast to her knowledge, she need not form the belief that she made a mistake, and, as a consequence she can continue to know, in the future, what she knows now. This suggestion might sound plausible at first, but it paves the way to an implausible form of dogmatism.[30] This is because this line of reasoning can be generalized to an ever greater number of consecutive drawings of a red marble. After all, although statistically unlikely, a sequence of one thousand, two thousand, ten thousand (or more) drawings of a red marble are also possible. But once the infallibilist sees this, she has no principled reason to stop here; there's nothing stopping her from endorsing a general principle that says this: 'Since I know that *p* now, I am now justified in considering any future evidence against *p* to be misleading, and, as such, dismissible.'[31] Of course, an allegiance to any such principle regarding how one should handle misleading evidence is equivalent to a rejection of defeasibilism. It is a rejection of the idea proposed by Harman in the previous section that one loses

[29] Many (for example, Brown 2018) claim that infallibilism is plagued with several *other* problems (e.g., skepticism). In that context, the dilemma I discuss here can be taken to be *yet another* stumbling block for the infallibilist.

[30] Maria Lasonen–Aarnio calls knowledge held in the face of counterevidence 'unreasonable knowledge' in Lasonen-Aarnio (2010).

[31] See Kripke (2011), and Borges (2015) for discussion of this form of dogmatism.

one's knowledge that *p* because one loses one's knowledge that new evidence against *p* is misleading. The cunning infallibilist we are imagining is not so easily tricked. She knows not to expose herself to misleading evidence. She has read her *Oddyssey*, and learned the Odyssean lesson. 'My resolve to avoid misleading evidence *now* will prevent my epistemic shipwreck *later*' she whispers to herself; 'because I *now* know that *p* I must tie *p* down and wax my ears lest I am mislead by the siren call of evidence against *p*.'

However plausible at first, this form of infallibilism ultimately leads to highly implausible consequences. It holds that Liz may rationally dismiss a sequence of ten thousand drawings of a red marble as misleading; it says that I can rationally dismiss the sheriff's phone call saying he found my car in the bottom of a lake as misleading evidence my car is not in the parking lot; and so on. And what is more, since it is rational for me to handle misleading evidence in this way, I can rationally reject evidence against what I know without even double-checking to make sure the proposition in question is true (why should I?). I take this type of result to be absurd. Of course I should double-check if my car is in the parking lot before I dismiss the sheriff's call. Similarly, it seems absurd to think that Liz should not double-check the contents of her bag before she dismisses as misleading evidence a sequence of ten thousand drawings of a red marble. The same applies, *mutatis mutandis*, to all similar cases.[32] So, horn (a) in the above dilemma is not ultimately viable, and the infallibilist should stay away from it.

Sadly, the infallibilist who accepts option (b) does not fare much better. To see why, remember, first, what is at issue in the threshold problem: vagueness. In particular, it seems implausible to think that a small difference in the factors that determine justification can make a big difference to whether one knows or not (again, this is structurally similar to the implausible suggestion that one becomes clearly bald by losing a single hair). Second, note also that option (b) embraces defeasibilism, since according to this view Liz satisfies the threshold for justification at *t* but not at *t**. Call this view, endorsed by Williamson and others, 'non-dogmatic infallibilism.' Non-dogmatic infallibilism and fallibilism converge on the case involving drawings with replacement. At *t*, Liz's justification satisfies the required threshold, *n*. At time *t**, after two hundred and fifty draws of a red marble, Liz's justification is defeated, and it falls below *n*. However, if x consecutive drawings of a red marble are enough to defeat Liz's justification, why not $x-1$ consecutive drawings of a red marble? Why would one single drawing make such a big epistemic difference? This, of course, is the diachronic threshold problem again. It is now clear that this problem also afflicts a popular version of infallibilism, non-dogmatic infallibilism. Because fallibilism and non-dogmatic infallibilism converge on Liz's case (and cases like it), they also converge on the diachronic form of the

[32] But some cases are not like the ones I describe here. In some cases the dogmatic stance seems rational. As Kripke noted (2011, p.49), most of us can rationally ignore any evidence suggesting that astrology, or necromancy amounts to an accurate description of reality. Kripke also notes that delineating when the dogmatic strategy is rational and when it is not is itself an epistemological problem. Nevertheless, the point in the body of text still stands—dogmatism cannot reasonably be applied to all cases where misleading evidence is a factor.

threshold problem. Non-dogmatic infallibilism and fallibilism converge in another interesting way too. Jettisoning defeasibilim is no more an option for the non-dogmatic infallibilist than it was for the fallibilist, since doing so pushes the infallibilist against the first horn of our dilemma—i.e., dogmatism.[33]

In sum, when we consider the effect of new misleading evidence on what one knows, it becomes apparent that the infallibilist is either impaled by the diachronic threshold problem or by dogmatism. The point of my discussion was to bring this difficulty to light. Hopefully, my contribution will elicit some discussion of this heretofore ignored issue. I will not argue for any particular response to this issue.

I will now turn to a few objections that might be raised against what I said in this section.

Objection: 'There is no diachronic threshold problem for infallibilism; Liz was never justified in believing that the bag contains one red marble and one black marble, so she did not cease to be justified after having drawn a red marble two hundred and fifty times.'

Reply: If this objection is correct, we have no reason to think that we are ever justified in believing anything. Liz *saw* the red marble and the black marble; she *knows* that the bag is a regular bag, that it has no abnormalities. Unless one is already convinced that infallibilism is true, one will find that it is more plausible to say that Liz is justified in believing that there is one red marble and one black marble in the bag before she starts drawing marbles from the bag. The claim that Liz is not justified in those circumstances is certainly consistent with infallibilism. Still, that does not mean that this claim is reasonable in a way that is not dependent on infallibilism. Moreover, note that this infallibilist interpretation of the marble case amounts to a rejection of defeasibilism, and we showed above that infallibilist views that reject defeasibilism fall pray to the dogmatism problem. The upshot is that the current objection gives us no reason to think that the diachronic problem is not a real problem for infallibilism.

Objection: 'Memory is what is causing the problem in the diachronic version of the marbles case. The same problem does not emerge in a case where memory is not involved. So, the diachronic threshold problem is actually a problem about memory knowledge, not knowledge as a whole.'[34]

Reply: As it turns out one can generate the same issue in cases where memory plays no significant role. Consider. Liz receives a commemorative quarter from the United States Mint. The coin commemorates the four hundredth birthday of philosopher and mathematician Blaise Pascal. As a fitting homage to the French philosopher of probability, the United States Mint produced a certificate for the coins. The certificate truthfully states that the coins are the fairest coins ever produced by mankind. After carefully reading the certificate and placing it right in front of her, Liz proceeds to flip the coin one thousand times. She observes the coin coming up heads every single time. If we suppose that Liz never loses sight of the

[33] In Borges (2020b), I explore another way in which this problem might be epistemologically interesting: I discuss whether the diachronic threshold problem upsets views that accept different Lockean theses connecting credences and (full-blown) belief. I argue that it does. Here, however, I ignore the issue.

[34] My thanks to Cesar Schirmer for raising this issue in conversation.

certificate, and that she takes copious notes of every single coin toss, then, I submit, she knows that her coin is fair *before* the thousandth toss, but fails to know that proposition *after* that toss.[35] Although Liz is looking at the certificate, she also knows that a big enough number of those coins were made and that she might have received a biased, defective coin. She knows that even though the process used by the United States Mint is highly reliable, it is not infallible. The relevant point here is, however, that Liz did not have to rely on her memory in any significant way in the process that lead her from knowledge to ignorance. Moreover, we can raise about this case what is essentially the same question we raised about the marbles case: why should one single coin toss make such a huge difference?

Objection: 'The diachronic problem arises because you are using discrete numerical values to describe degrees of justification. But given that justification is a qualitative notion, it cannot be represented by discrete, quantitative notions.'

Reply: the diachronic threshold problem arises for qualitative notions as well. Why should drawing a red marble (a small change in Liz's evidence) change the fact that Liz is fully/sufficiently justified in believing that there is a red and a black marble in the bag (a big epistemic change)? Numbers help make the problem obvious, they don't create the problem (the same applies to the synchronic version of the problem).

Objection: 'The diachronic threshold problem is simply a restatement of Williamson's anti-luminosity argument according to which no non-trivial mental state is luminous, where a non-trivial mental state, s, is luminous just in case one is in a position to know whether one is in s whenever one is, in fact, in s.'[36]

Reply: Not quite. Williamson's anti-luminosity argument is primarily about agents not being able to tell, from a first-person perspective, whether they are in state s or not. On the other hand, the diachronic threshold problem is about whether one *is* in the relevant state (i.e., knowledge) to begin with. In that sense, our problem logically precedes the luminosity issue.

5 Conclusion

I close with a few general lessons I believe we must draw from what came before. First, if the traditional, synchronic, threshold problem poses a real threat to fallibilism, then so does the diachronic version of the problem. Second, the most popular version of infallibilism (i.e., non-dogmatic infallibilism) also suffers from the diachronic threshold problem. Thirdly, by embracing defeasibilism, infallibilism loses some of the comparative advantages it claims to have over fallibilism.

I can already hear the fallibilist raising her voice and proclaiming to the infallibilist, in an accusatory tone, 'Tu quoque!'

[35] Again, if you find that one thousand tosses are not enough to move your intuitive needle, adjust accordingly. Also, as it should be obvious from the context of the case, I am assuming that the coin is in fact fair.

[36] Luis Rosa posed this question in conversation.

Acknowledgements This paper owes much to the input and encouragement of others. Stephen Hetherington kindly sent me detailed comments on an earlier draft. I discussed different versions of this paper with my colleagues from the University of Florida in The Epistemology Working Group: John Biro, Greg Ray, Chris Dorst, Jim Gillespie, and James Simpson. I presented some of the material in this paper at the 2019 meeting of the Southeastern Epistemology Conference in Greenville, North Carolina. I am thankful to that audience for their feedback; especially Mike Veber, Ted Poston, Andrew Moon, Jonathan Matheson, Kevin McCain, Matthew Frise and Elijah Chudnoff. I also presented a draft of this paper to the brown bag meeting led by Luis Rosa at the University of Cologne. My thanks to those who came to that meeting, in particular to Luis Rosa, Cesar Schirmer, Eve Kitsik and Francesco Praolini. I am also thankful for the helpful suggestions I received from two referees for *Philosophical Studies*. Lastly, this paper and the research that lead to it has been deeply influenced by many fruitful conversations over many years with Peter D. Klein and Claudio de Almeida. I am deeply grateful for their advice and friendship. Needless to say, I am fully and uniquely responsible for any mistakes that might remain in this paper.

Declarations

Conflict of interest The author declares that he has no confiict of interest.

References

Achinstein, P. (2003). *The book of evidence*. Oxford University Press.
Bacon, A. (2020). Inductive knowledge. *Nous, 54*(1), 354–388.
Bonjour, L. (2010). The myth of knowledge. *Philosophical Perspectives, 24*(1), 57–83.
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* (pp. 273–291). Oxford University Press.
Borges, R. (2020a). Knowledge from knowledge. *American Philosophical Quarterly, 57*(3), 283–297.
Borges, R. (2020b). *On Bayesian knowledge*. Draft.
Brown, J. (2014). Impurism, practical reasoning and the threshold problem. *Noûs, 48*(1), 179–192.
Brown, J. (2018). *Fallibilism: Evidence and knowledge*. Oxford University Press.
Chisholm, R. (1966). *Theory of Knowledge* (2nd ed.). Prentice Hall.
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 (pp. 292–311). Oxford University Press.
Descartes, R. (2008). Meditations on first philosophy: With selections from the objections and replies (M. Moriarty, Ed.). Oxford University Press.
Door, C., Goodman, J., & Hawthorne, J. (2014). Knowing against the odds. *Philosophical Studies, 170*(2), 278–287.
Dougherty, T. (2011). Fallibilism. In S. Bernecker & D. Pritchard (Eds.), *The Routledge companion to epistemology* (pp. 131–143). Routledge.
Dretske, F. (1971). Conclusive reasons. *Australasian Journal of Philosophy, 49*(1), 1–22.
Dretske, F. (1981). The pragmatic dimension of knowledge. *Philosophical Studies, 40*(3), 363–378.
Gendler, T. S., & Hawthorne, J. (2005). The real guide to fake barns. *Philosophical Studies, 124,* 331–352.
Hannon, M. (2017). A solution to knowledge's threshold problem. *Philosophical Studies, 174,* 607–629.
Hannon, M. (2019). *What's the point of knowledge?* Oxford University Press.
Harman, G. (1973). *Thought*. Princeton University Press.
Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford University Press.
Hetherington, S. (2006). Knowledge's boundary problem. *Synthese, 150,* 41–56.
Hintikka, J. (1962). *Knowledge and belief*. Cornell University Press.
Klein, P. (2019). How to get certain knowledge from fallible justification. *Episteme, 16,* 395–412. https://doi.org/10.1017/epi.2019.33.
Klein, P. D. (2008). Useful false beliefs. In Q. Smith (Ed.), *New essays in epistemology* (pp. 25–63). Oxford University Press.

Kotzen, M. (2019). A formal account of epistemic defeat. In B. Fitelson, R. Borges, & C. Braden (Eds.), *Themes from Klein*. Springer.

Kripke, S. (2011). Two paradoxes of knowledge. In *Philosophical troubles, collected Papers vol. I*. Oxford University Press.

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

Lehrer, K., & Thomas Paxson, J. (1969). Knowledge: Undefeated justified true belief. *The Journal of Philosophy, 66*(8), 225–237.

Malcolm, N. (1952). Knowledge and belief. *Mind, 61*(242), 178–189.

Neta, R. (2009). Williamson on knowledge. In P. Greenough & D. Pritchard (Eds.), (pp. 161–182). Oxford University Press.

Pollock, J., & Cruz, J. (1999). *Contemporary theories of knowledge*. Rowman and Littlefield.

Pritchard, D. (2016). *Epistemic angst: Radical skepticism and the groundlessness of our believing*. Princeton University Press.

Shogenji, T. (2012). The degree of epistemic justification and the conjunction fallacy. *Synthese, 184*(1), 29–48.

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

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

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.