

First-Class and Coach-Class Knowledge

Research Article

Spencer Paulson (ORCID: **0000-0001-5794-1387**)

Received: 10/2/2022

Revised: 12/2/2022

Accepted: 12/20/2022

Abstract: I will discuss a variety of cases such that the subject's believing truly is somewhat of an accident, but less so than in a Gettier case. In each case, this is because her reasons are not ultimately undefeated full stop, but they are ultimately undefeated with certain qualifications. For example, the subject's reasons might be ultimately defeated considered in themselves but ultimately undefeated considered as a proper part of an inference to the best explanation that is undefeated without qualification. In each section of the paper, I consider different qualifications and show how in each case we get an epistemic standing I call "coach-class knowledge". First-class knowledge requires justifying reasons that are undefeated without qualification. Coach-class knowledge only requires qualified lack of defeat. I will use this distinction to bring debates over knowledge from falsehood and fake barns to an ecumenical resolution. In both cases, the subject enjoys coach-class rather than first-class knowledge. I will also show that the defeasible reasoning tradition can better account for graded accidental truth than safety theories (Word Count: 11,892).

Keywords: Knowledge; Graded Knowledge; Defeasible Reasoning; Knowledge from Falsehood; Gettier problem; Fake Barns

Corresponding Author: Spencer Paulson ([spencerpaulson2023@u.northwestern.edu](mailto:spencerpaulson2023@u.northwestern.edu))

Running Title: "First-Class"

Here is a familiar idea from the defeasible reasoning literature: your reasons<sup>1</sup> need to pass muster along two dimensions of assessment for you to have knowledge. The first dimension requires that your *prima facie* reasons not be defeated by your total evidence. Otherwise, your belief is irrational or unjustified. But they also can't be defeated by facts of which you are unaware. Factual defeat can't make the belief irrational or threaten justification. However, it can threaten knowledge. If your justification is factually defeated, then it is (at best) accidentally true, no matter how rationally held. This is the upshot of Gettier cases. Call the first, rationality-involving dimension of assessment "subject-relativized" and the other fact-involving one "non-relativized". There is a fact of the matter about whether a subject's reasons are good along either dimension, the fact is just determined by different things in each case. It is determined by the subject's total evidence in the former and the totality of the facts in the second.

If this is right, then we can understand why Gettiered subjects have accidentally true beliefs. Non-accidentally true belief requires that the subject's reasons pass muster along both dimensions, but Gettier subjects have reasons that only do well along the former. This proposal is familiar.<sup>2</sup> I will argue here that it can also be used to explain graded accidentality. That is, some beliefs are neither straightforwardly knowledge nor non-knowledge. This is not (in the cases I'm thinking of) because "knowledge" is vague. Rather, it is because the subject's reasons pass along the first dimension of assessment full-stop but the latter dimension only with certain qualifications. Distinguishing different qualifications, then, allows the discovery of different ways that non-accidentally true belief

---

<sup>1</sup> I will generally talk about reasons since I am a reasons-firster. However, the point I want to make here can be put in terms of evidence if you prefer (cf. Klein 2017). Some locutions (e.g., "total evidence") sound better when I speak of evidence instead of reasons. The difference is only stylistic.

<sup>2</sup> See Lehrer (1965, 1970), Lehrer & Paxson (1968), Hilpinen (1971), Johnsen (1974), Swain (1974), Barker (1976), Klein (1971, 1976, 1980), Pollock (1986), Moser (1989), Audi (1993), Schroeder (2015), de Almeida & Fett (2015). See Shope (1983) for a useful overview.

can be graded.<sup>3</sup> I will argue that in each case we have instances of “coach-class” rather than “first-class” knowledge.

Accidental truth comes in degrees. First-class knowledge is completely non-accidentally true. It is the limiting point on the accidentality spectrum. Coach-class knowledge includes all cases where there is some accidentality involved, but not enough to make it a Gettier case. So, coach-class knowledge admits of differences in degree. There are also differences of kind among cases of coach-class knowledge. You get coach-class knowledge when your reasons are ultimately good relative to certain qualifications but not full-stop. Different qualifications give rise to different species of coach-class knowledge. Indeed, we will see a different species of the coach-class genus in each section. All species of the genus share a structural property: the subject’s justifying reasons are ultimately good insofar as x but not full stop. Different values of x give us different species. Although I will stick with the “first-class/coach-class” terminology, we could just as well call it “unqualified vs. qualified” knowledge.

Apart from providing an interesting new kind of case, this paper will also give us an ecumenical way of settling the disputes about knowledge from falsehood and fake barns. In these cases, we can split the difference in a principled way between the parties to the dispute. In the final section of the paper, I will briefly argue that a safety-based analysis of graded non-accidental truth doesn’t fare as well as the account I recommend. I leave open the possibility that a virtue-theoretic account can handle the cases of interest, although I won’t go into the matter here.<sup>4</sup>

In section (I) I will briefly go over the aspects of the defeasible reasoning tradition that matter for my proposal. In section (II) I will consider a new kind of case in which the subject’s

---

<sup>3</sup> The way I grade knowledge is distinct from, but compatible with, Reed’s (2013) way of doing it.

<sup>4</sup> See Carter (2014) for what I take to be the most promising virtue-theoretic approach to a similar phenomenon.

reasons are not ultimately good considered in themselves, but they are a proper part of the ultimately good reasons in the non-relativized sense. Insofar as we consider them as a proper part of those reasons, they are ultimately good. In section (III) I do the same thing except by focusing on cases where the subject's reasons are epistemically similar to the ultimately good reasons. I argue that cases of knowledge from falsehood are like this. In section (IV) I argue that intuitions about fake barns are less firm than those about classic Gettier cases because we must qualify our answer to the question whether their justifying reasons are ultimately good by distinguishing their foreground reasons from their background assumptions. The former are ultimately good while the latter are not. In section (V) I respond to objections and give reasons to choose the account developed here over a counterfactual analysis of graded non-accidentally true belief.

## **(I) Background**

In this section I will describe the core commitments of the defeasible reasoning tradition before going on to apply it in the following sections.<sup>5</sup>

A necessary condition for a subject to know *p* is that they believe *p* on the basis of a *prima facie* good reason. *Prima facie* reasons are foundational reasons. This doesn't mean that beliefs formed on their basis are infallible or incorrigible. A moderate foundationalist will hold that *prima facie* reasons don't depend positively on other reasons for support, but nonetheless depend negatively on the absence of countervailing reasons.<sup>6</sup> These countervailing reasons are called "defeaters". Defeaters are the mirror image of *prima facie* reasons. *Prima facie* reasons count provisionally in favor of believing something and defeaters count provisionally against believing it. For example, perceptual

---

<sup>5</sup> Some prominent objections to this approach are those of Feldman (2003), Foley (2012) and Turri (2012). In my view, an ample response to all three can be found in de Almeida & Fett (2015). Lasonen-Aarnio (2010 a,b, 2014) and Baker-Hytch & Benton (2016) raise a distinct set of worries I won't be able to address in this paper.

<sup>6</sup> The first person to put it this way was Robert Audi (1993), although the idea is latent in earlier work on defeasible reasoning.

experience as of a red wall is a *prima facie* reason to believe that the wall is red. It is an all-things-considered reason to so believe just in case it is undefeated. If the subject were to be told that the lighting in the room is non-standard, for example, they would have a defeater. Nothing I say in this paper will depend on any particular analysis of defeaters. This is a virtue, not a lacuna. The intuitive idea is that they are *prima facie* reasons to give up a belief (cf. Ballantyne 2015). There are many ways of trying to make the intuitive idea more precise, but the details don't matter for my purposes here.<sup>7</sup> The things I say here are independently plausible and a complete account of defeaters should strive to accommodate them.

Since defeaters are *prima facie* rather than all-things-considered reasons to give up a belief, your *prima facie* reason can have a defeater but still be a good reason all-things-considered. This can happen when the defeater is itself defeated. For example, if you were to learn that the person who told you the lighting is non-standard is a pathological liar, then the defeater would be defeated and your *prima facie* reason would be restored. Not all defeater-defeaters do this, however. Some defeater-defeaters give you new reasons rather than restoring your (actual) justifying reasons.<sup>8</sup> For instance, suppose you were to learn that the lighting on the wall really is non-standard but you were also told that someone painted the wall red so it would still look the same even if the lighting were to change. The information about the painting of the wall defeats the defeater having to do with the lighting. However, it doesn't restore your justifying reason (i.e., your perceptual experience). Despite the new information, your eyes are still not to be trusted. Your perceptual experience is now replaced, rather than restored, by the testimony about the painting of the wall. Going forward, I will call defeater-

---

<sup>7</sup> See for Pollock (1986) for one influential account. For recent work, see Brown & Simion (2021).

<sup>8</sup> Cf. Klein (1980), Pollock (1986), de Almeida & Fett (2015).

defeaters that do not restore the subject's justifying reasons "non-restoring defeaters"<sup>9</sup> and those that do restore their justifying reasons "restoring defeaters".

When we want to know whether a subject's belief is justified, we check to see if she has *prima facie* reason for that belief. If so, we check for defeaters in her possessed evidence. Non-restoring defeaters in her possessed evidence defeat her justification, so I will call them "justificational defeaters". But a belief can be justified yet fail to be knowledge. This is what we see in Gettier cases. In these cases, the subject's justification is defeated by a fact of which she is unaware. So, I will call these "factual defeaters".<sup>10</sup> The key insight of the defeasible reasoning tradition is that knowledge, over and above true belief, is a matter of the standing of the subject's reasons and to determine their standing we must look at those reasons from the subject's perspective but also *sub specie aeternitatis*. The subject's reasons must be good in a subject-relativized sense which requires that her justifying reasons not have a non-restoring defeater in her possessed evidence. It also requires that they be good in a non-relativized sense which consists in them not having a non-restoring factual defeater.

Peter Unger (1968) defined knowledge as non-accidentally true belief. Everyone seems to agree to this much, although there is much disagreement about what it is for a belief to be non-accidentally true. Some deny that we can understand the kind of accidents that undermine knowledge without a prior understanding of knowledge itself. These people are "knowledge-firsters".<sup>11</sup> Others think we can give a modal analysis of the kind of accidents that undermine knowledge that will help explain knowledge.<sup>12</sup> This is what safety theorists propose. Those working

---

<sup>9</sup> I get the term from de Almeida & Fett (2015). Similar terminology can be found in Klein (1980).

<sup>10</sup> Some call them "knowledge defeaters" (Audi 1993) or "propositional defeaters" (Bergmann 2006).

<sup>11</sup> See Williamson (2000), Sutton (2007), Bird (2007), Reynolds (2013). Some combine knowledge-first and virtue epistemologies, e.g., Miracchi (2015), Kelp (2016), Millar (2019).

<sup>12</sup> The version of this approach I will give the most attention is Pritchard (2005). See also Hawthorne (2004), Pritchard (2012) and Hawthorne & Rabinowitz (2017) for different ways of pursuing it.

in the defeasible reasoning tradition, alternatively, think we can understand the kind of accident that undermines knowledge in terms of *prima facie* reasons and defeaters.<sup>13</sup>

Importantly, both dimensions of epistemic assessment are explanatorily prior to knowledge and neither requires a modal analysis. The key components of the account are *prima facie* reasons and defeaters. *Prima facie* reasons are epistemic foundations. Proponents of the defeasible reasoning tradition are typically internalists about justification, which is the view that I recommend. I won't try to give a definition of internalism here, as there may be more than one internalism/externalism controversy. At the very least, however, internalists are committed to explaining the epistemic potency of the foundations in non-modal terms. There are a few options here. The epistemic potency might not be further explicable<sup>14</sup>, it might be explained in terms of a basic entitlement to believe that things are as they seem<sup>15</sup>, it might be explained in terms of the acquaintance relation<sup>16</sup>, or perhaps some other way. I will remain neutral here. Defeat isn't given a modal analysis either. Defeaters are either part of the subject's possessed evidence or actual world facts.

Now that we know how the defeasible reasoning tradition explains non-accidentally true belief, I will go on to discuss how it can be extended to explain graded non-accidentally true belief. In the final section, I will argue that it does better than a modal account could. A quick terminological note: when I talk of reasons being ultimately good/bad ultimately defeated/undefeated in this paper, I mean in the non-relativized sense of those terms, unless otherwise noted.

## **(II) Noir-Style Cases**

---

<sup>13</sup> See footnote 2 for representatives of this tradition. Klein (2017) calls these "quality of evidence" views.

<sup>14</sup> See Pryor (2001).

<sup>15</sup> See Huemer (2001, 2007).

<sup>16</sup> See Fumerton (1995).

The first kind of case I will consider I am going to call “noir-style cases” for lack of a better term. The basic idea is familiar enough from old movies and contemporary network police procedurals. There is a detective investigating something. They come up with a hypothesis about what happened. They later come across countervailing evidence suggesting that their initial theory was incorrect. After that, they come across even more evidence suggesting that their initial theory was not so much mistaken as incomplete. The true account of what happened ends up being a more convoluted version of what they initially thought. Raymond Chandler’s *The Big Sleep* patterns this way, although the plot is too complicated to discuss it here. Here is an outline of a simpler case,

**The Big Noir:** Marlowe was investigating the murder of James Cagney. Cagney was a known criminal and associate of crime boss John Huston. After asking around, Marlowe learned that Cagney was seen on more than one occasion leaving a hotel room with Huston’s wife Ingrid Bergman. At the end of the first day on the case, Marlowe formed the belief that Huston killed Cagney. He figured that Bergman and Cagney were having an affair and Huston killed Cagney when he learned about it. His reason was that this gives a cohesive explanation of the available evidence. The next day while looking around the hotel room Marlowe learned that it was used for a counterfeit operation. After asking around some more, he learned that Huston had been defrauded of money he had tied up in a counterfeit operation. Over the next couple of days other associates of Huston started appearing in dumpsters all over the city. Marlowe began to doubt his initial hypothesis. Kirk Douglas, Huston’s highest-ranking associate, then turned himself in and confessed to Marlowe that he, Cagney, Bergman and others had tried to defraud Huston of the money he had tied up in the counterfeit operation. During this time Bergman and Cagney began having an affair. Huston followed them to the hotel when he began to suspect Bergman of infidelity. When his suspicion of infidelity was confirmed, he also learned about the counterfeit scam. He



killed Cagney because of all he had learned and then went after all his associates that he suspected were in on the scam.

Marlowe's initial reason for thinking Huston did it was that the jealous husband narrative gave a cohesive explanation of the available evidence. Once more evidence came in, the explanation became less cohesive. Was the counterfeit somehow unrelated to the affair despite taking place in the same room? Were the killings of the other associates unrelated to the Cagney killing despite the fact that both were involved in the counterfeit scam? Once the last of the evidence came in the form of the confession, a new explanation became available. Marlowe's original explanation was a proper part of it. Insofar as Marlowe's original explanation is a part of the final explanation of the evidence, it offers an ultimately good reason for thinking Huston did it. Considered as a complete account of what happened, it does not offer an ultimately good reason for thinking that. Considered as a complete account it is defeated by the evidence it does not explain.

When Marlowe formed the belief that Huston did it at the end of the first day, there were plenty of factual defeaters. They did not defeat his justification because they were not part of his possessed evidence. They are like the defeaters we find in Gettier cases in this respect. However, it is less clear in this case than a classic Gettier case that Marlowe lacks knowledge. I submit that at the end of the first day on the case, his belief that Huston did it was somewhat non-accidentally true. This is because his reasons are ultimately good, subject to certain qualifications. We feel uneasy saying that they are ultimately good, but we also feel uneasy saying that they are not. They aren't as straightforwardly ultimately defeated as the Gettier victim's reasons. They are also not as straightforwardly ultimately good as the reasons a subject has when we have no qualms about attributing knowledge to them.

Notice that this is not the already discussed phenomenon of partial defeat.<sup>17</sup> Justification comes in degrees and a defeater can defeat all of it or part of it. A partial defeater lowers the degree of justification the subject enjoys but without lowering it so much that it is no longer sufficient to support knowledge.

What we have in **The Big Noir** is not a reason that is defeated just enough to put the subject's level of justification right at the threshold for knowledge when it was previously much higher than that. This is not why we are uneasy about attributing knowledge to Marlowe early in the story. Marlowe's reasons give him far less than is needed for knowledge when they are considered as a complete account of what happened. There is simply too much that his IBE doesn't explain. But when we think of those same reasons as a proper part of the explanation that ultimately emerges, they are ultimately good considered in that light.<sup>18</sup>

Something similar happens when a scientist's justification for believing that some event will occur is that their theory predicts it, later investigation reveals defeaters for their theory while yet further investigation reveals that their original theory is a special case of the correct one. It is similar in the following respect. Considered as a complete account, their original theory is not an ultimately good reason for believing the event will take place. Considered as a proper part of the true theory, it is an ultimately good reason. If a Newtonian scientist (for example) believes truly that some event will occur because their theory predicts it and their theory is a special case of the true theory of that

---

<sup>17</sup> See Thune (2010) for more discussion.

<sup>18</sup> It could be objected that when Marlowe reasons this way, there is an implicit "that's all there is to it" assumption or further belief that he has captured that core of the matter. If so, then his reasons are fully defeated (by the rebutting defeater that there is more to it). I respond that his assumption that this is all there is to it is fully defeated but that doesn't mean he lacks coach-class knowledge. Indeed, this case is analogous to cases where a scientist of the past has a theory that is strictly speaking false but correct about some domain they are drawing inferences about. In these cases, I urge the subject's reasons aren't fully defeated even though they think their theory is a complete account of the target domain rather than just a special case of the correct theory. Thanks to Sandy Goldberg for raising this concern.

target domain, is their belief accidentally true? Not exactly, even if the theory on which their prediction is based is thoroughly discredited by evidence they do not possess. The theory is thoroughly discredited as a complete account of physical objects generally because it is unable to explain phenomena within that domain. Nonetheless, there is a sense in which the theory is true and consequently a sense in which the Newtonian's reasons for believing the prediction are ultimately good. Insofar as we restrict our attention to the sense in which the theory is true and the reasons are ultimately good, the Newtonian's belief strikes us as non-accidentally true. If we don't draw the distinction between the two senses in which reasons can be or fail to be ultimately good, then we just feel uneasy saying either that the reasons are or are not ultimately good.

The source of our uneasiness is that Marlowe and the Newtonian have coach-class knowledge. For them to have first-class knowledge, their reasons would have to be ultimately good *simpliciter* rather than ultimately good insofar as this but not insofar as that. Nonetheless, they have a derivative kind of knowledge precisely because there is a derivative sense in which their reasons are ultimately good. Their reasons have a humble form of ultimate goodness because of how they relate to the reasons that are ultimately good *simpliciter*. The proper parthood relation is a little different in each case.<sup>19</sup> In the former, it is just one conjunct of a conjunctive explanation. In the latter, the scientist inadvertently treats a special case of the correct theory as if it were a complete account. The result is that they unwittingly avail themselves of a useful model. In either case though, the subject has a somewhat but not entirely accidentally true belief because their reasons were a proper part of the reasons that yield first-class knowledge.

Of course, inference to the best explanation is just one form justification might take and the reader might worry that IBE is idiosyncratic in many respects. You might worry that the things I say

---

<sup>19</sup> Thanks to Tez Clark for pointing out the need to be clear about this.

about it don't generalize to other forms of justification. We will see in sections (II) and (III) that the point generalizes to knowledge that doesn't depend on IBE. What IBE shares with these other kinds of justification is susceptibility to both unqualified and qualified defeat. This is what makes room for the first-class/coach-class distinction.<sup>20</sup>

## (II) Noir Variations

In this section I will consider a variation of the Noir-style case in which the subject's reasons are not a proper part or a special case of the ultimately good reasons but are relevantly similar. I will then apply the lessons learned to the knowledge from falsehoods debate.

Suppose we alter **The Big Noir** so that Marlowe mistook drug residue for counterfeit residue. Suppose further that he comes up with the same account he ends up with at the end of the original case but with no help from Kirk Douglas. That is, he believes that Huston killed Cagney because he learned about the counterfeit operation and the affair. His reason is an inference to the best explanation. He was right, except that it was a drug operation.

Marlowe's reason is ultimately defeated, at least in the strictest sense. Was his belief accidentally true? I feel uncomfortable saying this without qualification. I also feel uncomfortable saying it was non-accidentally true without qualifying my answer. The key is in the fact that his explanation is defeated *in the strictest sense*. Marlowe's reasons are ultimately defeated in the strictest sense, but they are nonetheless pretty much the same as the ultimately undefeated reasons.<sup>21</sup> Marlowe's explanation resembles the ultimately good one enough for us to want to qualify our statement that his reasons are ultimately defeated. We want to say that he was on the right track, he

---

<sup>20</sup> Thanks to an anonymous referee for pointing out the need to be clear about this.

<sup>21</sup> Cf. Klein (2008).

was basically right or something of the sort. For this reason, the truth of his belief doesn't strike us as a total accident.

Some might ask if Marlowe's justifying reason is really that there is illegal activity of some sort, rather than a belief about the nature of that activity specifically.<sup>22</sup> If so, his justifying reason is not ultimately defeated. Though it is easy enough to imagine a case like this, we are free to stipulate that the case we are dealing with isn't. For whatever reason, let us assume that Marlowe doesn't have a more general belief than his belief about the drug operation or at least that he doesn't base anything on it.<sup>23</sup> If you are still skeptical, consider the following case: I see a cougar and form the belief that I am in danger before I draw any inferences about being confronted by a member of the Puma genus, a large cat, etc. I still know I am in danger. I submit that I have coach-class knowledge if we change the case so that I have a mistaken belief that it is a jaguar from which I infer that I am in danger.

Since I don't form a more general belief about the cat *before* inferring that I am in danger, my belief that I am in danger can't be based on it. Knowledge entails doxastic justification and a reason for believing p can only doxastically justify belief that p if the belief that p is based on it. Since my belief that I am in danger is not based on any mental state having to do with the size or genus of the cat, my doxastic justification (and consequently knowledge) can't depend on justifying reasons having to do with these things. Granted, I am disposed to believe that there is a large cat as soon as I see it. But a *disposition to believe* can't enter into the basing relation. You can't base a belief off a disposition. A *dispositional belief*, on the other hand, could be the basis for another belief. A dispositional belief requires that I have actually stored the content believed, it is dispositional

---

<sup>22</sup> Cf. Ball & Blome-Tillman (2014), Montminy (2014) for two different ways of developing this thought.

<sup>23</sup> Cf. Coffman (2008), Klein (2008).

because I am not entertaining that belief in an occurrent episode, although I am disposed to do so.<sup>24</sup> However, in the case we are imagining I stipulate that I haven't yet processed and stored the content PUMA or LARGE CAT but rather only the more specific content COUGAR. So, I don't have a dispositional belief (as opposed to a disposition to believe) that I am confronted by a large cat or a puma. The burden of proof is on anyone who wants to dispute the psychological plausibility of my stipulation. The burden will be difficult to discharge since it is ultimately an empirical question and work on feature detection, both in AI and mammalian brains, suggests that we first detect low-level features and then proceed bottom-up to increasingly abstract ones.<sup>25</sup>

The Cougar example is just like the noir variation in the relevant respects. The subject forms a justified false belief with a specific content and draws an inference from it before forming a more general true belief that could have, but didn't, appear in his reasoning as a premise. Marlowe's only reason is ultimately defeated, at least in the strictest sense. The question is once again whether he knows. My suggestion is that it is harder to say one way or the other than it is in obvious cases of knowledge or non-knowledge. This is because his reasons are not ultimately good considered in themselves but are ultimately good insofar as they resemble reasons that are ultimately good *simpliciter*. If we just ask whether his reasons are ultimately good or whether he knows without drawing the in-themselves/insofar as they are similar to x-distinction, we will have a hard time giving an unequivocal answer.

I think something along these lines is what is going on in cases of knowledge from falsehood. It was originally thought that one's justification for a true belief depending on a false lemma (that one justifiably believes) was necessary and sufficient for gettierization (Clark 1963). It was later discovered that one's justification could be defeated without any false lemmas (Lehrer &

---

<sup>24</sup> For more on this distinction, see Audi (1994).

<sup>25</sup> See Gauker (2011), Khaligh-Razavi (2014), Yamins & Di Carlo (2016) & Buckner (2018).

Paxson 1968; Goldman 1976). New evidence might compromise the connection between one's justification and the belief it is supposed to justify without showing the former to consist of any falsehoods. If so, then perhaps false lemmas are sufficient but not necessary for gettierization. However, the sufficiency claim has also been contested. Some cases seem to involve false lemmas in such a way that they don't compromise the subject's knowledge. Consider a case in which a professor believes she has 50 students in her class, so she prints 60 handouts for her lecture just to be safe. She then forms the belief that she has printed enough handouts because she has 60 and she only needs 50. However, she was wrong about having 50 students in her class. She actually has 51. She miscounted.<sup>26</sup> If you worry that she must have also had a more general true belief that there are approximately 50 students because otherwise she would have had no reason to print more than 50 handouts, we can adjust the case so that her TA prints 60 handouts. This way we don't have to explain why she printed 60. When the TA hands them off to her immediately before the lecture, she sees that there are 60 handouts (5 stacks of 12, say), reasons that there are 50 students and that 60 is greater than 50. For these reasons (and only these reasons), she infers that there are enough.<sup>27</sup>

Some argue that false lemmas don't always preclude knowledge.<sup>28</sup> Others take the hardline stance that a false lemma always precludes knowledge.<sup>29</sup> I recommend a compromise. The professor's reasons are not ultimately good in the strictest sense, since they depend on her ignorance of the actual number of students in her class. If she weren't ignorant, her justifying reasons would be

---

<sup>26</sup> This is an adaptation of a case from Warfield (2005).

<sup>27</sup> Thanks to an anonymous referee for bringing this issue to my attention.

<sup>28</sup> See Saunders & Champawat (1964), Hilpinen (1988), Klein (1996: 106, 2008), Hawthorne (2004: 57), Warfield (2005), Coffman (2008), Fitelson (2010, 2017), Feit & Cullison (2011), Arnold (2013), Hiller (2013), de Almeida (2017), Hawthorne & Rabinowitz (2017), Buford & Cloos (2018), Turri (2019), Luzzi (2019), Zhao (Forthcoming).

<sup>29</sup> See Aristotle *Posterior Analytics* 71a2-3, Russell (1912: 76), Clark (1963), Armstrong (1973: 198-9), Harman (1973: 47, 120, 1980), Lehrer (1974: 220), Swinburne (2001: 199), Feldman (2002: 36-7), Audi (2003), Lycan (2006: 156-7), Williamson (2007: 145-7), Kripke (2011: 202), Littlejohn (2013), Ball & Blome-Tillman (2014), Schnee (2014), Montminy (2014), Kelp (2016), Borges (2017) and Yong Lee (2021).

different, since they are in fact defeated.<sup>30</sup> So, she doesn't have knowledge in the strictest sense. Nonetheless, after remedying her ignorance, her justifying reasons wouldn't have to change much. So, there is some inclination to say that the falsehood is benign, just as there is an inclination to say that she doesn't know in the strictest sense. Why not say both? Why not say that she doesn't know *in the strictest sense*, but she nonetheless enjoys coach-class knowledge on account of her reasons being similar enough to the ultimately good reasons?

It is important to note that the subject's reasons don't just need to be semantically similar to the ultimately good ones.<sup>31</sup> They need to be similar in the right way for the subject to enjoy coach-class knowledge. To bring this point out, consider the following Gettier case. Marlowe believes that Huston is guilty because a generally reliable informant told him that the guilty party was wearing a blue jacket on the day of the crime and the pullman porter told him earlier that Huston was wearing a blue jacket on the day of the crime. It turns out the informant just made that up so he could sell Marlowe phony intel. It just so happens the guilty party was wearing a green jacket on the day of the crime. Huston did it and the pullman porter saw him entering the train afterward, but it turns out (unbeknownst to all) the pullman porter is colorblind.

This is a Gettier case rather than a case of coach-class knowledge. The detective believed truly that Huston did it. His reasons were the testimony of the informant and the porter regarding jacket color. However, Huston wasn't wearing a blue jacket. He was wearing a green jacket. That he was wearing a green jacket is an ultimately good reason to think he was the guilty party. Green is similar to blue though, just as 51 is similar to 50. So, why don't we have a case of coach-class knowledge, as we did in the case of the handouts?

---

<sup>30</sup> Cf. Montminy (2014).

<sup>31</sup> Thanks to Sandy Goldberg for pointing this out.



Although the detective's reason is semantically similar to an ultimately good one, it is not similar epistemically.

**Epistemic Similarity:** A good (in the relativized sense) reason to believe  $p$ ,  $R$ , is epistemically similar to a good (in the non-relativized sense) reason to believe  $p$ ,  $N$ , just in case  $R$  and  $N$  are semantically similar and the way they support  $p$  is also similar.

Reasons are semantically individuated, so semantic similarity matters. But what interests us here is not just their semantic properties. We are also interested in what it is about them that makes them capable of supporting belief that  $P$ . Once I apply the epistemic similarity principle to the cases of interest, I will say a bit more about the principle itself and how limited its ambitions are.

Here is how this account of epistemic similarity helps us distinguish the gettiered Marlowe case from the professor with coach-class knowledge. In the former case, the detective's reason for believing Huston did it is the match between two sources of testimony regarding jacket color. The informant told him that the offender was wearing a blue jacket and the pullman porter told him that Huston, one of his suspects, was wearing a blue jacket that same day. These are Marlowe's justifying reasons. Those reasons are factually defeated. The informant was lying and the porter is colorblind. There is a semantically similar undefeated reason: Huston was wearing a green jacket and the culprit was wearing a green jacket. Despite the similarity of green to blue, the undefeated reason doesn't support belief that Huston did it in the same way. Marlowe's justification came from received testimony and the undefeated reasons have nothing to do with that testimony.

In the case of the professor things are different. Her reasons are semantically similar to the ultimately good ones in that 51 is similar to 50. Furthermore, her reasons work in the same way as the ultimately good ones. Both work by positing more handouts than students. This gives the professor a *prima facie* reason to think she has enough. That reason is defeated on account of her

being wrong about the exact number. Part of her justifying reason is that there are exactly 50 students in the class, that part of her justifying reason is rebutted by the fact that she missed one student. However, a semantically similar reason that supports the subject's belief in the same way (i.e., by positing a favorable handout/student ratio) is undefeated.

I should note that, in general, similarity is pragmatically squishy and somewhat elusive. I don't expect necessary and sufficient conditions for exactly the right amount of similarity to be forthcoming. Any attempt to use counterfactual analyses in terms of a similarity relation on possible worlds will be the same in this respect. We can talk about all or nearly all nearby worlds but that still leaves open exactly how many and how nearby they must be. Furthermore, different epistemologists might disagree about the details regarding the way a subject's justifying reasons support a particular belief. My proposal here isn't supposed to help us adjudicate those disputes. Rather, it tells us how to account for grades of knowledge once we have adjudicated them.

Some readers may not feel the pull of the intuition that the professor doesn't know in the strictest sense. Her belief was safe, and this may be what drives the intuition. To make the point even more forcefully, suppose she made 1,000 handouts instead of 60 but was still wrong (by the same margin) about the number of students. Do we still want to say she doesn't know in the strictest sense?

I think we should. Knowledge is non-accidentally true belief. If knowledge is non-accidentally true belief, then somewhat accidentally true belief is somewhat knowledge. We should all agree that the professor's belief is more accidentally true than it would be if she had a true belief about the number of people in the audience. You should grant that even if you endorse some other

account of accidentalness, such as the safety account.<sup>32</sup> I will give reasons to prefer my account to the safety account in the final section.

### (III) Post-Gettier Cases

In this section I hope to show that so-called “post-Gettier cases<sup>33</sup>” often fit into the pattern that is the theme of this paper. We should first consider Ginet’s Potemkin barn case.<sup>34</sup> S is driving through rural Wisconsin and sees what appears to be a barn on the roadside. Lighting conditions are normal, her vision is good, etc. So, she believes that she is looking at a barn. However, unbeknownst to her, she is in a county that has met with economic hardship. The denizens of this once prosperous county put up barn facades so the economic downturn is less apparent to those passing by. It just so happens that S was looking at the only real barn left in Potemkin barn county.

Does she know it is a barn? The most common answer is “no”. It seems that the truth of her belief is accidental, at least to many.<sup>35</sup> Another response is that her perceptual equipment is working properly and her belief was formed on the basis of it, so there is no problem.<sup>36</sup>

Perhaps predictably at this point, I offer a compromise. S enjoys coach-class but not first-class knowledge that she is looking at a barn.<sup>37</sup> To make this move, I must identify a qualified sense

---

<sup>32</sup> Cf. Carter (2014) who reaches a similar conclusion by a very different argument.

<sup>33</sup> That is, cases that feel like Gettier cases but involving non-inferential knowledge (so no false lemmas). I get the term from Peter Graham (2000).

<sup>34</sup> From Goldman (1976).

<sup>35</sup> Some of whom are Goldman (1976), Williamson (2000), Howard-Snyder, Howard-Snyder & Feit (2003), Pritchard (2005, 2012), Lackey (2006, 2009), Reed (2009), Greco (2010, 2012), Luzzi (2010), Kelp (2013), Carter (2014), Goldberg (2015, 2018, 2019).

<sup>36</sup> Some of whom are Brandom (1994), Hetherington (1999), Hawthorne & Gendler (2005), Lycan (2006), Sosa (2007), Turri (2011), Feit & Cullison (2011).

<sup>37</sup> Sosa says the fake barn subject has animal knowledge but not reflective knowledge (2007), which is a different way of trying to split the difference. I have worries about going his way, some of them are about the tenability of the animal/reflective distinction as applied to cognitively mature humans (see Goldberg & Matheson 2020). Even if those worries can be addressed, I still hope to show that reflective knowledge is graded and account for fake barns in that way.

in which her reasons are ultimately good, even though they aren't ultimately good *simpliciter*. My thought here is that her foreground reason is ultimately good though it is functioning against background assumptions that are not. Her foreground reason is her perceptual evidence. Perceptual evidence only gives one reason to believe that things are as they appear so long as one is making certain background assumptions, e.g., that one's perceptual system is working properly, lighting conditions are standard, etc. It is epistemically ok for the subject to assume these things. Otherwise, it wouldn't feel like a Gettier case at all but rather an unjustified true belief case. Nonetheless, there is a (factual) defeater for one of her background assumptions. The defeater is that she is in Potemkin barn county where there are many barn facades indiscernible from real barns when viewed from the highway. It is not clear which background assumption is defeated. It does seem clear, however, that she was assuming that conditions weren't as they in fact were. Perceptual evidence gives one justification for believing that things are as they appear, other things equal. This is a clear case in which other things are not equal, though S had no reason to suspect this.

This means that one of her background assumptions has a defeater. This by itself doesn't preclude knowledge. If the defeater has a defeater-defeater, then the defeater could be misleading and her justifying reason could remain intact. There is some inclination to think that this is what is going on here. The fact that she is in Potemkin barn county is a misleading undercutting defeater because she is looking at the only real barn in the county. So, her perceptual equipment is trustworthy in the case in which she finds herself after all. It only seems to be untrustworthy if we consider her perceptual equipment relative to Potemkin barn county in general, but there is a logically stronger reference class relevant here: the particular barn she is staring at. This particular situation is free of the perils with which the rest of Potemkin barn county is fraught. So, the thought is that once we realize which part of Potemkin barn county she is in, the fact that she is in Potemkin

barn county no longer gives us reason to doubt her senses. That is, the defeater is defeated in such a way that her original perceptual justification is restored.

I suspect this sort of thinking is part of what (at least tacitly) drives the intuition that she knows. Her perceptual justification is restored, after all. Nonetheless, the truth of her belief still seems somewhat coincidental. Perhaps this isn't a problem. It is generally agreed that it can be a coincidence that you have the evidence you do without that undermining knowledge.<sup>38</sup> However, the problem is deeper than that. It isn't just that she was lucky to get the evidence she did. Her evidence only has evidential force relative to certain background assumptions. It has evidential force relative to the background assumptions she permissibly makes (i.e., that these are normal conditions) and it has evidential force relative to the background facts of which she is unaware (i.e., this is the only spot in the county where appearances aren't misleading). Her belief is accidentally true because the background assumptions she is actually making are defeated. There are other background assumptions relative to which her perceptual justification would have evidential force, but these are not the ones she is actually making and it wouldn't be permissible for her to make them anyway. I don't have an account of what makes it permissible to assume things in the background. What seems clear is that despite not explicitly thinking about it, we make some background assumptions and not others at any given time.<sup>39</sup> Furthermore, some background assumptions are epistemically impermissible, at least in the sense that they can't support justified belief or knowledge (e.g., assuming you are in Potemkin barn county under completely normal circumstances). Similarly, others are permissible in the sense that they can support justified belief and knowledge.

Furthermore, background assumptions are no less susceptible to defeat than beliefs are. Whatever

---

<sup>38</sup> Cf. Nozick's (1981) masked bandit case.

<sup>39</sup> Cf. Bach (1985).

the correct account of permissible background assumptions turns out to be, it will need to accommodate these things.

The upshot is that insofar as her (foreground) reasons are considered in themselves they are undefeated and the subject's belief is non-accidentally true. However, when they are considered as part of a package that includes the background assumptions that enable them to acquire epistemic potency, they are defeated because the assumptions on which they depend are defeated.

Fake barns differ from standard Gettier cases in the following way. Consider the Ferrari case from Lehrer (1965). Mr. Nogot is a student in S's class. Mr. Nogot tells S that he (Nogot) owns a Ferrari. S sees Nogot driving it in to campus one day. This perceptual evidence and the testimony from Nogot give S reason to think that Nogot owns a Ferrari. However, unbeknownst to S, Nogot is a pathological liar and he was borrowing the Ferrari from someone else in S's class. S has no reason to suspect this, so she forms the belief that Nogot has a Ferrari and then existentially generalizes to the belief that someone in her class does, which is true but not knowledge.

S's perceptual evidence is factually defeated by the fact that Nogot was borrowing the car. S's testimonial evidence is defeated by the fact that Nogot is a pathological liar. The perceptual evidence is only perceptual evidence that Nogot owns a Ferrari against the background assumption that people generally own the cars they drive and this situation is not unusual. The testimonial evidence is only evidence that Nogot owns a Ferrari against the background assumption that testimony about car-ownership is generally reliable and this situation is not unusual. The situation is, however, unusual in such a way that neither background assumption really applies to her situation. The background assumptions that this situation is usual are both factually defeated. In this respect, the case is just like the fake barn case. However, unlike in the fake barn case, there aren't any facts of which S is unaware that could function as background assumptions relative to which her testimonial

and perceptual justification could re-gain traction, were she to learn more about her circumstances. Consider her testimonial justification. Testimony is generally reliable, but Nogot belongs to the smaller and logically stronger reference class of pathological liars. So, his testimony doesn't give one reason to believe what he says. Unlike the fake barn case, Nogot's testimony on this occasion doesn't belong to some yet smaller reference class such that his membership in that class could enable the testimonial justification to re-gain traction. He is a pathological liar and this is one of the cases where he was indeed lying. This is unlike the fake barn case in which the circumstances in the county are generally misleading but the subject finds herself in a sub-region of it where they are not misleading.

In the classic Gettier case, there is something wrong with the subject's foreground reasons because there are no (undefeated) background assumptions that a better-informed subject could make and relative to which the same foreground reasons could justify the subject's belief. This differs from the "post-Gettier" fake barn case where the foreground reasons, as such, are fine. They just can't support knowledge when the subject is making the background assumptions she in fact is.

#### **(IV) Objections, Replies & Follow-up Discussion**

I will start by addressing a worry people sometimes have when I propose this view. It might appear that first-class knowledge is incredibly difficult to come by if my view is true, as the world is bound to be rife with defeaters that aren't part of any given subject's possessed evidence.<sup>40</sup>

There are a few things worth saying here. The first is that even if this is true, it wouldn't be that bad. It would only show that real cases of knowledge are rarely paradigmatic. My claim is that we get our grip on what knowledge is by first understanding what it is for a belief to be non-

---

<sup>40</sup> Thanks to Jennifer Lackey and Sandy Goldberg for pressing me on this one.

accidentally true. This should be relatively (though not entirely<sup>41</sup>) uncontroversial. As I've said above, I recommend understanding non-accidentally true belief in the way familiar from the defeasible reasoning tradition. I go beyond that tradition in showing how the core ideas of it make room for graded accidentality. So long as it is granted that the subject's believing truly is somewhat accidental and that paradigmatic cases of knowledge are non-accidentally true full stop, the objector grants me all I need. The paradigmatic cases (i.e., first-class knowledge) give us our grip on non-accidentally true belief and, consequently, knowledge. So, they are explanatorily prior to the messier cases of coach-class knowledge. The latter only count as knowledge to the extent that they approximate the paradigmatic cases that give us our grip on the concept of knowledge in the first place.

If the objector is worried about anything, it should be that others in the defeasible reasoning tradition are committed to an implausible form of skepticism. But I mitigate that skepticism by admitting gradations of knowledge. If my view is a kind of skepticism, then those of Klein (1980), Pollock (1987) and Moser (1989) are all more thoroughgoing varieties of skepticism, though the worry about skepticism doesn't tend to arise regarding these views.<sup>42</sup> None of these authors put any constraints on factual defeaters. Since that is the feature of my account that arouses the worry about skepticism, it is just as much a worry about their accounts as mine.

Some might object along the following lines: Justification is graded and knowledge entails justification, but it doesn't follow that knowledge is graded. Anything above a certain threshold suffices for knowledge. Similarly, one could object that just because accidentality is graded and knowledge is non-accidentally true belief, it doesn't follow that knowledge is graded. Perhaps

---

<sup>41</sup> Knowledge-firsters will of course object. I can't address their concerns here.

<sup>42</sup> Though see Swain (1974) for a preemptive objection to the kind of view shared by these authors calling it "an epistemologist's pipe dream", which sounds like a charge of skepticism to my ears.



anything beyond a certain threshold for non-accidentality is knowledge *tout court*. If so, then perhaps there is no first-class/coach-class distinction.<sup>43</sup>

If we go this way, then we will have to get by with two categories (i.e., knowledge and Gettier cases) where I think we need three (i.e., first-class knowledge, coach-class knowledge and Gettier cases). The problem is that the taxonomy is too coarse, and it will require us to group dissimilar phenomena together. If we shrink the taxonomy to two categories, then we either have to group the cases I have been considering in this paper with Gettier cases or we have to group them with paradigmatic cases of knowledge.

Let us first consider the costs of grouping the cases of interest with Gettier cases. The problem is that they are explanatorily dissimilar. In classic Gettier cases, the subject's reasons are defeated. In the cases of interest here, the subject's reasons are also factually defeated, but they are factually undefeated relative to certain qualifications. Factually undefeated reasons are what separate mere JTB from K. The most natural way of accommodating the point about qualified lack of defeat is that it results in qualified knowledge. In general, if there are jointly sufficient conditions for a property, then meeting one of those conditions relative to certain qualifications will result in having that property relative to those qualifications. Being a male and being unmarried are jointly sufficient for being a bachelor. If you are unmarried, but only relative to certain qualifications, then you are a bachelor relative to those qualifications. If you were the groom in a marriage performed within a religious organization not recognized by the state, then you are a bachelor relative to certain qualifications. We might say you are a "census bachelor". One could insist that knowledge is an exception to this general rule, but that requires argument. You need to tell us what it is about the

---

<sup>43</sup> Thanks to an anonymous referee for raising this concern.

concept of knowledge that makes it unsusceptible to qualification. Perhaps this can be done, but the burden is on you to show that it can.

Let us now consider the other option: grouping the cases of interest with first-class knowledge. This approach runs into a different version of the same problem. Once again, it requires that we group dissimilar phenomena together. Knowledge is non-accidentally true belief. To understand knowledge, we have to understand what it is for a belief to be non-accidentally true. We understand non-accidental truth in the strictest sense first. That is, first-class knowledge comes first in the order of explanation. We then come to understand qualified non-accidental truth as an approximation of first-class knowledge. This shows us that first-class and coach-class knowledge need to be distinguished by the theorist.

One could of course grant the claim about the relation of explanatory priority between two types of case but deny that this results in a distinction between grades of knowledge. You might insist that at least some instances of what I am calling “coach-class knowledge” are close enough to the “first-class” end of the spectrum to count as knowledge *simpliciter*.

I grant that coach-class knowledge is quite often close enough *for practical purposes*. Two phenomena might be importantly different for the theorist but similar enough so that we often don't have to distinguish them in practice. My claim is that the epistemologist ought to distinguish the two, even if the layperson can get by without the distinction most of the time. The fact that people typically don't draw this distinction is not evidence that the distinction isn't real or philosophically important. Compare: people used to think Jadeite and Nephrite were the same mineral. Specialists eventually discovered that they were wrong: despite looking the same to the untrained eye, the two differ in molecular structure. I claim that first-class knowledge and coach-class knowledge are like Jadeite and Nephrite in that they are different, but people are generally unaware of the difference.

Because people are generally unaware of the difference, it is not lexically marked in ordinary language. It would consequently be a mistake to try to determine whether there is a difference by looking at felicity tests and other linguistic data.<sup>44</sup> In short: we need some reason to think the explanatory difference between two kinds of case doesn't track a conceptual difference between two kinds of knowledge. The burden is on those opposed to the distinction to show that there is no conceptual difference.

I have argued that either way of trying to collapse the first-class/coach-class distinction comes with a burden of proof. This is not the only reason to retain the distinction. The first-class/coach-class distinction enables us to resolve long standing disputes in epistemology ecumenically. Rather than dismissing one side as simply confused or stubborn, we have excavated insights from both sides of the debates about fake barns and false lemmas. The people who deny that the subjects in such cases have knowledge are right about something important: the difference between (for example) fake barn cases and paradigmatic cases of knowledge. The people who claim that the subjects know are also right about something important: the difference between fake barns and classic Gettier cases. Philosophical work is typically not without difficulties, but it is rarely completely misguided. I have developed a framework for explaining why this is true in a couple of well-known debates in epistemology.

The next thing to consider are the costs and benefits of adopting this view as opposed to a safety<sup>45</sup> account of accidental truth. Safety theorists don't typically talk about degrees of safety. They typically talk about safety outright. This makes sense because they are just interested in whatever amount of it is needed for knowledge. Nonetheless, safety can clearly be graded. Possible worlds

---

<sup>44</sup> See Stanley (2005: Chapter 2) for an inquiry into whether knowledge is graded that proceeds by looking into linguistic data.

<sup>45</sup> I use "safety" here broadly enough to include safe-methods accounts such as those of Hawthorne (2004) and Pritchard (2012: 257). In cases where the differences might be relevant, I will discuss them in footnotes.

where you believe falsely using a similar belief forming process can be more or less remote. If a safety theorist were to try to make sense of graded accidental truth, then the obvious way to go would be to try to make sense of it in terms of degrees of safety. If safety is what makes for first-class knowledge, then presumably it is a lesser degree of safety that makes for coach-class knowledge.

One point worth briefly noting is that my proposal is compatible with Necessitarianism and the safety view is not. Necessitarianism is the view, endorsed by Spinoza, that there is only one logically possible world. Necessitarianism entails determinism but is not entailed by it. I am not arguing that Necessitarianism is true or even that we should spend much time worrying about it. Nonetheless, it still seems right to say that even if we were to learn (surprisingly) that it is true, we should still want to allow room for the possibility of accidentally true belief. When I imagine myself learning that Spinoza was right all along, I find myself reconsidering many of my philosophical views. But I don't find myself reconsidering whether there are Gettier cases. Yet the safety theorist would have no choice but to reconsider. Nobody with a true belief could have easily believed falsely since there are no nearby worlds at which they or their counterparts could believe falsely. Similarly, I would still want to be able to vindicate the sense that there is an epistemically significant distinction between coach-class knowledge and first-class knowledge even if there are no true counterfactuals and hence no way to draw the distinction in terms of them. This is not because I am sympathetic to Necessitarianism but rather because my commitment to the philosophical significance of these distinctions in epistemology seems orthogonal to my disagreement with Spinoza.

Somewhat relatedly, Jennifer Lackey and others<sup>46</sup> have offered Gettier cases such that the subject's belief is safe. Consider the following,

*Southernmost Barn:* While entering a Midwestern farming community on her cross-country drive, Janice looked at the first barn that she saw, which was on the southernmost end of the field, and formed the corresponding belief 'There is a barn'. As it happens, the barn she saw is the only real one, surrounded by barn façades that members of this community have placed in the field in order to make their town appear prosperous. However, as a matter of strict and unwavering policy, the members of this community always place their only real barn on the southernmost end of their land, since this is where traffic first enters their town. Moreover, thirty years earlier, Janice had lived in a house on the southernmost end of this field in the precise location of the one real barn. Because of her deep interest in her childhood roots combined with the brief period during which she can safely take her eyes off of her driving, she would invariably have looked at only the particular place in the field where the real barn exists. (Lackey 2006)

The problem is that the belief is made safe in a way that has no bearing on the quality of the subject's reasons. The subject's belief has a defeater, just as in any Gettier case. It is safe despite that defeater not being defeated in a way that restores the epistemic potency of the subject's justifying reason. This shows us that safety can be achieved without eliminating knowledge-undermining luck.

If we can make non-knowledge just as safe as knowledge, we can make coach-class knowledge just as safe as first-class knowledge in the same way. We just need to write vignettes with causally relevant facts that don't defeat the defeaters. In some cases, we don't even need to do that.

---

<sup>46</sup> Neta & Rohrbaugh (2004), Comesaña (2005), Hiller & Neta (2007), Kelp (2009), Coffman (2010), Bogardus (2014), Goldberg (2015). It matters for my purposes that at least one of these cases works, not that they all do.

Consider the case of the Newtonian scientist from earlier. They truly predict an event by deriving it from a true description of antecedent conditions and Newtonian laws. They falsely believe that Newtonian physics is a complete physical theory when it is in fact just a special case of a more general theory. I said earlier that this subject has coach-class knowledge but not first-class knowledge. The truth of her belief is more accidental than it would be if she were to know the correct physical theory, but it is less of an accident than a classic Gettier case. But the subject's methods are equally safe in either case. Either way, for them to believe falsely either the laws of physics or the antecedent circumstances would have to be different.<sup>47</sup> By focusing on the quality of the subject's reasons, rather than modal properties that are explanatorily orthogonal to them, I provide an account that focuses solely on the epistemically relevant factors.

Lastly, consider a variation of Warfield's handout cases but where the professor undercounts rather than overcounts. She has 60 handouts and believes that will be enough because she counted 50 students. However, she counted one student twice and she really has 49. Her belief that she has enough is even safer than it would be if she had counted correctly. More would need to change for her to believe falsely that she has enough, since the difference between students and handouts is greater. However, it seems like her belief that she has enough is epistemically worse off rather than better off since it is defeated.

## **(V) Conclusion**

---

<sup>47</sup> Pritchard (2005, 2007) restricts his safety-theory to knowledge of logically and nomically contingent propositions. The proposition known, in my example, is contingent although it is derived from nomic necessities, so it isn't clear that Pritchard's restriction helps here. Furthermore, it doesn't seem that the safe-methods approach he adopts later (Pritchard 2012: 257) meant to deal with necessary truths is going to help here. The methods of the Newtonian and the Einsteinian are equally safe when applied to the domain where their theories are empirically equivalent and this is the context in which both subjects find themselves.

I hope to have shown that knowledge admits of gradations, not because justification does, but because accidental truth does. There are several ways this can happen. I discuss a different one in each section and there are likely to be others I haven't considered. This is significant in part because it allows us to bring the debates about knowledge from falsehood and fake barns to ecumenical resolutions. Furthermore, I hope to have shown that the defeasible reasoning tradition can give a better account of what is going on in these cases than a modal account.

**Acknowledgements:** I would like to thank Tez Clark, Sandy Goldberg, Regina Hurley, Jennifer Lackey, Nate Lauffer, Whitney Lilly, Kathryn Pogin and Baron Reed for helpful comments.

#### **References:**

- Armstrong, D.M.** 1973. *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.
- Arnold, A.** 2013. 'Some Evidence is False.' *Australasian Journal of Philosophy* 91(1): 165-72.
- Audi, R.** 1993. *The Structure of Justification*. Cambridge: Cambridge University Press.
- 1994. 'Dispositional Beliefs and Dispositions to Believe'. *Noûs* 28 (4):419-34.
- 2003. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*, Second Edition. London: Routledge.
- Bach, K.** 1985. 'A Rationale for Reliabilism.' *The Monist* 68: 246-63.
- Baker-Hytech, M. & Benton, M.** 2015. 'Defeatism Defeated.' *Philosophical Perspectives* 29 (1): 40-66.
- Ball, B. & Blome-Tillman, M.** 2014. 'Counter Closure and Knowledge Despite Falsehood.' *Philosophical Quarterly* 64(257): 552-68.
- Ballantyne, N.** 2015. 'The Significance of Unpossessed Evidence.' *Philosophical Quarterly* 65 (260): 315-35.
- Barker, J.** 1976. 'What You Don't Know Won't Hurt You?' *American Philosophical Quarterly* 13:303-8.
- Bergman, M.** 2006. *Justification without Awareness: A Defense of Epistemic Externalism*, New York: Oxford University Press.
- Bogardus, T.** 2014. 'Knowledge Under Threat.' *Philosophy and Phenomenological Research* Vol. LXXXVII, No. 2. 289-313.
- Borges, R.** 2017. 'Inferential Knowledge and the Gettier Conjecture.' In. Borges, R., de Almeida, C., Klein, P. (eds.) *Explaining Knowledge: New Essays on the Gettier Problem*. Oxford: Oxford University Press. Pp. 292-311.
- Brandom, R.** 1994. *Making it Explicit*. Cambridge, MA: Harvard University Press.

- Brown, J. & Simion, M.** 2021. *Reasons, Justification and Defeat*. New York: Oxford University Press.
- Buckner, C.** 2018. 'Empiricism Without Magic: Transformational Abstraction in Deep Convolutional Neural Networks'. *Synthese*, 195, 5339-5372.
- Buford, C. & Cloos, C.M.** 2018. 'A Dilemma for the Knowledge Despite Falsehood Strategy.' *Episteme* 15(2): 166-82.
- Carter, J.A.** 2014. 'Robust Virtue Epistemology as Anti-Luck Epistemology: A New Solution.' *Pacific Philosophical Quarterly* 97: 140-55.
- Clark, M.** 1963. 'Knowledge and Grounds: A Comment on Mr. Gettier's Paper.' *Analysis* 24 (2):46-8.
- Coffman, E.J.** 2008. 'Warrant Without Truth.' *Synthese* 162(2): 173-94.
- 2010. 'Misleading Dispositions and the Value of Knowledge.' *Journal of Philosophical Research* 35: 241-58.
- Comesaña, J.** 2005. 'Unsafe Knowledge.' *Synthese* 146: 395–404.
- de Almeida, C.** 2017. 'Knowledge, Benign Falsehoods and the Gettier Problem.' In: Borges, R., de Almeida, C., Klein, P. (eds.) *Explaining Knowledge: New Essays on the Gettier Problem*. Oxford: Oxford University Press. Pp. 292-311.
- de Almeida, C. & Fett, J.R.** 2015. 'Defeasibility & Gettierization: A Reminder.' *Australasian Journal of Philosophy* 94(1): 152-69.
- Feit, N. & Cullison, A.** 2011. 'When Does Falsehood Preclude Knowledge?' *Pacific Philosophical Quarterly* 92:283-304.
- Feldman, R.** 2003. *Epistemology*. Upper Saddle River, NJ: Pearson.
- Fitelson, B.** 2010. 'Strengthening the Case for Knowledge from False Belief.' *Analysis* 70(4):666-9.
- . 2017. 'Closure Counter-Closure and Inferential Knowledge.' In: Borges, R., de Almeida, C., Klein, P. (eds.) *Explaining Knowledge: New Essays on the Gettier Problem*. Oxford: Oxford University Press. Pp. 312-24.
- Foley, R.** 2012. *When Is True Belief Knowledge?* Princeton: Princeton University Press.
- Gauker, C.** 2011. 'Words and Images: An Essay on the Origin of Ideas.' New York: Oxford University Press.
- Gettier, E.** 1963. 'Is Justified True Belief Knowledge?' *Analysis* 23: 121-3.
- Goldberg, S.** 2015. 'Epistemic Entitlement and Luck.' *Philosophy and Phenomenological Research* Vol. XCI, No. 2: 273-302.
- . 2018. *To the Best of Our Knowledge: Social Expectations and Epistemic Normativity*. Oxford, UK: Oxford University Press.
- . 2019. 'A Normative Account of Epistemic Luck.' *Philosophical Issues* 29(1).
- Goldberg, S. & Matheson, J.** 2020. 'The Impossibility of Mere Animal Knowledge for Reflective Subjects.' *Erkenntnis* 85(4): 829-40.



- Goldman, A.** (1976) ‘Discrimination and Perceptual Knowledge.’ *The Journal of Philosophy* 73(20): 771-91.
- Graham, P.** 2000. ‘Transferring Knowledge.’ *Noûs* 34(1): 131-52.
- Greco, J.** 2010. *Achieving Knowledge: A Virtue Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- . 2012. ‘A (Different) Virtue Epistemology.’ *Philosophy and Phenomenological Research* Vol. 85, No. 1: 1-26.
- Harman, G.** 1973. *Thought*. Princeton, NJ: Princeton University Press.
- . 1980. ‘Reasoning and Explanatory Coherence.’ *American Philosophical Quarterly* 17(2):151-7.
- Hawthorne, J.** 2004. *Knowledge and Lotteries*. Oxford, UK: Oxford University Press.
- Hawthorne, J. & Gendler, T.** 2005. ‘Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies.’ *Philosophical Studies* 124(3): 331-52.
- Hawthorne, J. & Rabinowitz, D.** 2017. ‘Knowledge and False Belief.’ In. Borges, R., de Almeida, C., Klein, P. (eds.) *Explaining Knowledge: New Essays on the Gettier Problem*. Oxford: Oxford University Press. Pp. 325-44.
- Hetherington, S.** (1999) ‘Knowing Failably.’ *Journal of Philosophy* 96, pp. 565 – 587.
- Hiller, A.** (2013) ‘Knowledge Essentially Based upon False Belief.’ *Logos & Episteme* IV(1):7-19.
- Hiller, A. & Neta, R.** 2007. ‘Safety and Luck.’ *Synthese* 158(3), 303–13.
- Hilpinen, R.** 1971. ‘Knowledge & Justification.’ *Ajatus*. 33:7-39.
- 1988. ‘Knowledge and Conditionals.’ in Tomberlin, J. (ed.) *Philosophical Perspectives 2, Epistemology* (pp. 157-82). Atascadero, CA: Ridgeview.
- Howard-Snyder, D., Howard-Snyder, F. & Feit, N.** 2003. “Infallibilism and Gettier’s Legacy” *Philosophy & Phenomenological Research* 66: 304-27.
- Johnsen, B.** 1974. ‘Knowledge.’ *Philosophical Studies* 25:273-82.
- Kelp, C.** 2009. ‘Knowledge and Safety.’ *Journal of Philosophical Research* 34: 21-31.
- 2013. ‘Knowledge: The Safe-Apt View.’, *Australasian Journal of Philosophy*, 91:2, 265-278.
- 2016. ‘Justified Belief: Knowledge-First Style.’ *Philosophy and Phenomenological Research* Vol. XCIII No. 1. Pp. 79-100.
- Khaligh-Razavi, S.-M., & Kriegeskorte, N.** 2014. ‘Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation.’ *PLoS Computational Biology*, 10(11). <https://doi.org/10.1371/journal.pcbi.1003915>
- Klein, P.** 1971. ‘A Proposed Definition of Propositional Knowledge.’ *Journal of Philosophy* 68: 471-82.
- 1976. ‘Knowledge, Causality, Indefeasibility.’ *Journal of Philosophy* 76:792-812.
- 1980. ‘Misleading Evidence and the Restoration of Justification.’ *Philosophical Studies* 37:81-9.

- 1996. ‘Warrant, Proper Function, Reliabilism, and Defeasibility.’ In J. Kvanvig (Ed.), *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge* (pp. 97-130). Lanham, MD: Rowman & Littlefield.
  - 2008. ‘Useful False Beliefs.’ In Smith, Q. (Ed.), *Epistemology: New Essays*. Oxford: Oxford University Press, pp. 25-62.
  - 2017. ‘The Nature of Knowledge’. In Borges, R., de Almeida, C. & Klein, P. (eds.) *Explaining Knowledge: New Essays on the Gettier Problem*. Oxford: Oxford University Press, pp. 35-56.
- Kripke, S. 2011. *Philosophical Troubles, Volume 1*. Oxford: Oxford University Press.
- Lackey, J.** 2006. ‘Critical Study: Pritchard’s *Epistemic Luck*.’ *The Philosophical Quarterly* Vol. 56, No. 3: 284-9.
- 2009. ‘Knowledge and Credit.’ *Philosophical Studies* 142: 127-42.
- Lehrer, K.** 1965. ‘Knowledge, Truth and Evidence.’ *Analysis* 25: 168-75.
- 1970. ‘The Fourth Condition for Knowledge: A Defense.’ *Review of Metaphysics* 24:122-128.
  - 1974. *Knowledge*. Oxford: Clarendon Press.
- Lehrer, K. & Paxson, T.** 1968. ‘Knowledge: Undefeated Justified True Belief.’ *Journal of Philosophy* 66:225-37.
- Littlejohn, C.** 2012. ‘No Evidence is False.’ *Acta Analytica* 28: 145-59.
- Luzzi, F.** 2010. ‘Counter-Closure.’ *Australasian Journal of Philosophy* 88(4):673-83.
- 2019. *Knowledge from Non-Knowledge: Inference, Testimony and Memory*. Cambridge: Cambridge University Press.
- Lycan, W.** 2006. ‘On the Gettier Problem Problem.’ In Hetherington, S.C. (ed.) *Epistemology Futures* (pp. 148-68). New York: Oxford University Press.
- Millar, A.** 2019. *Knowing by Perceiving*. Oxford: Oxford University Press.
- Miracchi, L.** 2015. ‘Competence to Know.’ *Philosophical Studies*, 172: 29-56.
- Montminy, M.** 2014. ‘Knowledge Despite Falsehood.’ *Canadian Journal of Philosophy* Vol. 44 No. ¾.
- Moser, P. K.** 1989. *Knowledge & Evidence*. Cambridge: Cambridge University Press.
- Neta, R. & Rohrbaugh, G.** 2004. ‘Luminosity and the Safety of Knowledge.’ *Pacific Philosophical Quarterly* 85: 85(4): 396-406.
- Nozick, R.** 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Pollock, J.** 1986. *Contemporary Theories of Knowledge*. 1<sup>st</sup> Edition. Rowman & Littlefield: Totowa, NJ.
- Pollock, J. & Cruz, J.** 1999. *Contemporary Theories of Knowledge*. 2<sup>nd</sup> Edition. Lanham, ML: Rowman & Littlefield.
- Pritchard, D.** 2005. *Epistemic Luck*. Oxford: Oxford University Press.
- 2007. ‘Anti-Luck Epistemology.’ *Synthese*, 158, (3). 277-97.
  - 2012. ‘Anti-Luck Virtue Epistemology.’ *The Journal of Philosophy* 109(3): 247-279.

- Reed, B.** 2009. 'A New Argument for Skepticism.' *Philosophical Studies* 142: 91-104.
- 'Fallibilism, Epistemic Possibility and Epistemic Agency.' *Philosophical Issues*, 23: 40-69.
- Russell, B.** 1912. *The Problems of Philosophy*. Oxford: Oxford University Press.
- Saunders J.T. & Champawat N.** 1964. 'Mr. Clark's Definition of 'knowledge.' *Analysis* 25(1), 8-9.
- Schnee, I.** 2014. 'There is No Knowledge from Falsehood.' *Episteme* 12(1)53-74.
- Schroeder, M.** 2015. 'Knowledge is Belief for Sufficient (Objective and Subjective) Reason.' In *Oxford Studies in Epistemology*, volume 5, (eds.) Tamar Szabó Gendler and John Hawthorne. pp. 226-252.
- Shope, R.** 1983. *The Analysis of Knowing: A Decade of Research*. Princeton University Press.
- Sosa, E.** 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1*, Oxford, UK: Clarendon Press.
- Stanley, J.** 2005. *Knowledge and Practical Interests*. Oxford University Press: Oxford, UK.
- Swain, M.** 1974. 'Epistemic Defeasibility.' *American Philosophical Quarterly* 11:15-25.
- Swinburne, R.** 2001. *Epistemic Justification*. Oxford: Oxford University Press.
- Thune, M.** 2010. 'Partial Defeaters and the Epistemology of Disagreement.' *The Philosophical Quarterly* 60(239): 355-72.
- Turri, J.** 2011. 'Manifest Failure: The Gettier Problem Solved.' *Philosophers' Imprint* 11: 1-11.
- 2012. 'In Gettier's Wake.', in *Epistemology: The Key Thinkers*, ed. Stephen Hetherington, London: Continuum. pp.214-29.
- 2019. 'Knowledge from Falsehood: An Experimental Study.' *Thought: A Journal of Philosophy* 8(3):167-78.
- Unger, P.** 1968. 'An Analysis of Factual Knowledge.' *Journal of Philosophy* 65:157-70.
- Warfield, T.** 2005. 'Knowledge from Falsehood.' *Philosophical Perspectives* 19:405-16.
- Williamson, T.** 2000. *Knowledge and its Limits*. New York: Oxford University Press.
- 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.
- Yamins, D. L., & DiCarlo, J. J.** 2016. 'Using Goal-Driven Deep Learning Models to Understand Sensory Cortex.' *Nature Neuroscience*, 19(3), 356.
- Yong Lee, K.** 2021. 'Reconsidering Alleged Cases of Knowledge from Falsehood.' *Philosophical Investigations* 44(2):151-62.
- Zagzebski, L.** 1994. 'The Inescapability of Gettier Problems.' *Philosophical Quarterly* 44:65-73.
- Zhao, B.** (Forthcoming) 'Knowledge from Falsehood, Ignorance of Necessary Truths and Safety.' *Philosophia*.