

Warrant without truth?

E. J. Coffman

Received: 7 November 2006 / Accepted: 17 April 2007 / Published online: 13 July 2007
© Springer Science+Business Media B.V. 2007

Abstract This paper advances the debate over the question whether *false* beliefs may nevertheless have **warrant**, the property that yields knowledge when conjoined with true belief. The paper’s first main part—which spans Sections 2–4—assesses the best argument for **Warrant Infallibilism**, the view that only *true* beliefs can have warrant. I show that this argument’s key premise conflicts with an extremely plausible claim about warrant. Sections 5–6 constitute the paper’s second main part. Section 5 presents an overlooked puzzle about warrant, and uses that puzzle to generate a new argument for **Warrant Fallibilism**, the view that *false* beliefs can have warrant. Section 6 evaluates this pro-Fallibilism argument, finding ultimately that it defeats itself in a surprising way. I conclude that *neither* Infallibilism *nor* Fallibilism should now constrain theorizing about warrant.

Keywords Warrant · Knowledge · Infallibilism · Fallibilism

1 Introduction

Many true beliefs do not qualify as knowledge: knowledge requires more than mere true belief. Following current convention in epistemology, let **warrant** name the property that “makes the difference” or “fills the gap” between mere true belief and knowledge: warrant is the (possibly logically complex) property that yields knowledge when conjoined with true belief. This paper advances an important current debate about warrant. I will lay out the paper’s plan after locating and motivating the relevant debate.¹

¹ Michael **Huemer** (2005) has recently shown that those discussing questions like the ones raised here should be more careful when fixing the reference of ‘warrant’. Huemer argues that no feature is simply *the one* that yields knowledge when conjoined with true belief. Suppose there is some such feature, W. The fact that W exists entails that there is another property that yields knowledge when conjoined with true belief—viz., *being false or having W*. So, if there is a feature that yields knowledge when conjoined with true belief, then more than one feature does so: no property is *the one* that yields knowledge when conjoined with true belief.

Edmund Gettier's influential "Is Justified True Belief Knowledge?" (1963) has prompted philosophers to explore (at least) two related questions about warrant. The first—salient since Plato's *Meno* and *Theaetetus*—is this:

- What is the correct substantive theory of warrant?

The main lesson of Gettier's paper is that a prominent kind of **epistemic justification**—one expressed by locutions like 'S is justified in believing that p'—is not identical with warrant. Gettier conveys this lesson by describing true beliefs that are justified yet fall short of knowledge, where this failure is due to the presence of a kind of knowledge-precluding luck that justification cannot rule out all by itself. Gettier's examples thus impugn a traditional—not to mention natural—approach to warrant that identifies warrant with justification. To solve the so-called **Gettier Problem**, we must find a plausible condition on knowledge that excludes the kind of knowledge-precluding luck illustrated by Gettier-type cases. (Notably, Gettier leaves open the question whether warrant *requires* justification, as do I in what follows.)

A second question explored in the wake of Gettier's paper is this:

- What interesting adequacy constraints are there on theorizing about warrant?

Nowadays, many think Gettier-type cases teach not only that warrant is not identical with justification, but also that any theory which allows that *false* beliefs can have warrant is inadequate. The basic idea is this. The particular theory of warrant Gettier refuted allows that false beliefs can be justified *and* identifies warrant with justification. Gettier's target thus entails that a *false* belief may nevertheless have warrant. The fact that one such theory of warrant succumbs to Gettier-type cases raises the worry that *any* such theory will succumb to such cases. We will consider rigorous versions of this line of thought below.

Thus, many think that an adequate theory of warrant must entail that only *true* beliefs can have warrant. Sticking with current convention, let **Warrant Infallibilism** be the view that only *true* beliefs can have warrant, and let **Warrant Fallibilism** be the view that *false* beliefs also can have warrant.² To reiterate the last point in the terminology just introduced: many think that a theory of warrant is adequate only if the theory entails *Infallibilism*, only if the theory prohibits a false belief's having warrant.

Here is the question I will explore in what follows:

- Should one or the other of Infallibilism and Fallibilism constrain our theorizing about warrant?³

Footnote 1 continued

In light of this argument, how should one wishing to discuss questions like those raised here fix the reference of 'warrant'? I propose the following: **warrant** is the (possibly logically complex) *epistemically valuable* feature that yields knowledge when conjoined with true belief—where a property is **epistemically valuable** iff its being exemplified entails the obtaining of an epistemically good or positive state of affairs. Because some false beliefs have no epistemically valuable features—e.g., a false belief grounded entirely in wishful thinking—no property that has *being false* as a disjunct qualifies as epistemically valuable (in the relevant sense). So, while Huemer's argument establishes that the description typically used to fix the reference of 'warrant' is unsatisfied, his argument does not also show that the alternative reference-fixing description just suggested is unsatisfied.

² Mainly because there is significant precedent for doing so, I will here typically use 'Infallibilism' and 'Fallibilism' as names for these views. See, e.g., Merricks (1995) and Howard-Snyder et al. (2003).

³ *Pointed Question*: Does not the description by which you fixed the reference of 'warrant'—viz., 'the (possibly logically complex, epistemically valuable) property that yields knowledge when conjoined with true belief'—entail that only *true* beliefs can have warrant? *Answer*: No. Granted, the description tells us *only* what warrant

Much of the post-Gettier literature concerns the question what must be added to true belief to get knowledge. Many who take up this question commit themselves—at least implicitly—to a particular side in the Infallibilist–Fallibilist debate by advancing *either* some substantive theory of warrant *or* a “nonpartisan” argument for one of the views in question (i.e., an argument that does not assume some substantive theory of warrant).⁴ In this way, much of the post-Gettier literature can be seen as a debate between Warrant Infallibilists and Warrant Fallibilists. Besides being intrinsically interesting, this debate is important because each side excludes certain attractive conceptions of knowledge.⁵ On the one hand, if *Fallibilism* is right, then a belief’s being knowledge does not require that the belief be caused by something connected with its truth, or that there be no fact such that the belief would be unjustified were one to believe the fact. On the other hand, if *Infallibilism* is right, then knowledge requires more than true belief based on fitting, reliable evidence that is neither overridden nor undercut by other evidence one has. Thus, reflection on the Infallibilist–Fallibilist debate stands to help us steer clear of certain attractive—yet ultimately mistaken—conceptions of knowledge.

This paper advances the Infallibilist–Fallibilist debate by assessing *both* an improved version of the best argument for Infallibilism *and* a new nonpartisan argument for Fallibilism. Notably, this new argument for Fallibilism stems from an overlooked puzzle about warrant. Moreover, the argument seems to be the first of its kind in the literature: an extensive search of recent work in epistemology fails to uncover an attempt at a *nonpartisan* argument for Fallibilism.⁶

This paper has two main parts. The first—which spans Sects. 2–4—assesses two versions of the best argument for Infallibilism, what we will call the **Supervenience Argument**. Section 2 explains and evaluates the most prominent version of the Supervenience Argument, which is due to Trenton Merricks (1995, 1997).⁷ Section 3 develops a better version of the Supervenience Argument. Section 4 appraises this argument, finding ultimately that it fails: its key premise conflicts with an extremely plausible claim about warrant.

Sections 5–6 constitute the paper’s second main part. In Sect. 5, I introduce a puzzle about warrant, and use it to generate a new nonpartisan argument for Fallibilism—the **Egalitarianism/Preservationism Argument** (the **EPA**). Section 6 assesses the EPA, finding ultimately that it too fails: the argument defeats itself in a surprising way. I conclude that, contrary to what many think, *neither* Infallibilism *nor* Fallibilism should now constrain our theorizing about warrant.

One last preliminary. To set the stage for Sects. 2–4, I must narrow the pool of candidates for best pro-Infallibilism argument down to the Supervenience Argument. Why reckon the Supervenience Argument the best argument for Infallibilism? We can start this process by eliminating all *partisan* arguments from our pool of candidates: since every substantive theory of warrant is at least fairly controversial, no partisan argument is likely to provide strong

Footnote 3 continued

does when had by a true belief—viz., it makes the belief a case of knowledge. That is the kernel of truth in the pointed question. But the description does not also say that warrant can be had *only* by true beliefs.

⁴ For this usage of ‘nonpartisan’, see Howard-Snyder et al. (2003).

⁵ Cf. Zagzebski (1994), Merricks (1995) and Howard-Snyder et al. (2003). See Sect. 5.1 for more on this and the following two claims.

⁶ The fact that Infallibilism and Fallibilism seem roughly equal in initial plausibility further motivates this paper’s project. If Infallibilism and Fallibilism enjoy roughly equal initial plausibility, the only reasonable way to take a side in this debate is to endorse an *argument* for one of the views; and then the question whether there is a *good* argument for either view becomes more important than it might otherwise have been.

⁷ Arguments similar to Merricks’s are also suggested in work by Zagzebski (1994, 1996, 1999), Bonjour (1996), Klein (1996) and Bigelow (unpublished).

reason to believe Infallibilism. Turning our attention to *nonpartisan* arguments, then, an extensive review of the recent literature uncovers three,⁸ the best of which is the Supervenience Argument. I will close this section by briefly discussing the other two nonpartisan arguments for Infallibilism.⁹

One of these arguments claims that no condition sufficient to exclude the kind of knowledge-preventing luck illustrated by Gettier-type cases could be met by a *false* belief.¹⁰ To ensure that we have some handle on such cases, consider this famous one due to Lehrer (1965):

Nogot works in S's office. One day, Nogot provides S with lots of evidence that Nogot owns a Ferrari. Naturally enough, S comes to believe—on the basis of that evidence—that Nogot owns a Ferrari. S then infers that someone in his office owns a Ferrari. Now, Nogot has been shamming: he does not really own a Ferrari. As it happens, though, another of S's colleagues, Havit, just now bought a Ferrari. So, S has a justified true belief that someone in his office owns a Ferrari; but he does not know that someone in his office owns a Ferrari.

S has a kind of knowledge-precluding luck—call it **Gettier Luck**—with respect to his justified true belief that someone in his office owns a Ferrari. Again, the nonpartisan argument for Infallibilism now under consideration claims that no condition sufficient to exclude Gettier Luck could be met by a *false* belief. Since any adequate account of warrant must include such a condition, no such account could be met by a false belief: only *true* beliefs can have warrant.

This argument fails, for there are conditions sufficient to exclude Gettier Luck that may nevertheless be met by a false belief.¹¹ To see this, consider the requirement for warrant laid down by the following thesis:

Counterfactual Requirement (CR): A belief has warrant only if the belief would not be just **luckily true** for its subject were the belief true. (Here, 'luckily true' expresses a concept illustrated by paradigm Gettier-type cases. The sense of 'luckily true' in play is that in which, e.g., S's inferential belief that someone in his office owns a Ferrari is just luckily true—and so, is not a case of knowledge.)

Whereas CR's requirement on warrant *cannot* be met by a "Gettierized" belief,¹² that requirement *can* be met by a false belief.¹³ While asleep at home, I have an exceptionally vivid

⁸ Cf. Howard-Snyder et al. (2003).

⁹ For thorough critical discussion of these arguments, see Sects. 1–2 of Howard-Snyder et al. (2003).

¹⁰ For such an argument, see Zagzebski (1994, 1996, 1999).

¹¹ This criticism was developed by Howard-Snyder et al.; see Sect. 1 of their 2003.

¹² Any Gettierized belief is of course luckily true (in the relevant sense), and thus such that it might be luckily true were it true, and thus *not* such that it would be nonluckily true were it true. (Here and elsewhere, I make the standard assumption that 'would' and 'might' are interdefinable as follows: 'If p were the case then q would be the case' is true iff it is false that 'If p were the case then not-q might be the case'.)

¹³ Howard-Snyder et al. (2003: 310) offer an example to illustrate this claim. I think their example leaves something to be desired. Here is their case:

[A] false belief can be such that if it were true, it would not be accidentally true. Consider the following illustration of the point. We visit the zoo one Saturday morning and rush to our favorite display: the zebra... [W]e see what looks like a zebra. Naturally enough, we believe that *there's a zebra*. However, suppose that last night the zookeeper, Fred, inadvertently poisoned the zoo's only zebra, Zak, and in order to keep zoo-goers from being disappointed, painted his mule, Moses, to look exactly like Zak. Nevertheless, it is plausible to suppose that our belief would not be accidentally true if it were true. For although in fact there is no zebra before us, the closest worlds at which there is are worlds in

dream, and form the *false* belief that I am teaching on campus. If this false belief were instead *true* (i.e., were I teaching on campus, believing myself to be doing so), I would know that I am teaching on campus. So, if my false belief that I am teaching on campus were instead true, that belief would constitute knowledge: a false belief may nevertheless be such that it would not be luckily true were it true. Contrary to the argument under consideration, then, some conditions sufficient to exclude Gettier Luck may be met by false beliefs.¹⁴

The linchpin of the second of the two failed pro-Infallibilism arguments mentioned above is this:

- **If** there were a warranted false belief, **then** its warrant could be transferred via inference to a luckily true belief.¹⁵

The problem is that this claim is at best counterbalanced. For starters, the claim seems roughly equal in initial plausibility to Fallibilism, and thus cannot impugn the latter all by itself.¹⁶ Moreover, it is hard to see what might tip the scales in the relevant claim's favor. Notice, for instance, that the claim does not obviously enjoy support from so-called "closure" principles for justification and knowledge.¹⁷ For one thing, such principles do not obviously entail the claim: justification is not identical with warrant, and no false belief constitutes knowledge. For another thing, such principles are themselves questionable, and so are not likely to yield a strong argument for Infallibilism.¹⁸

So much, then, for the two failed pro-Infallibilism arguments just discussed. In the next three sections, I will thoroughly assess two versions of the remaining nonpartisan argument for Infallibilism—the Supervenience Argument.

Footnote 13 continued

which, for example, Fred doesn't poison Zak, or he finds an actual zebra to replace poor Zak instead of Moses the mule. In these worlds, we form the belief that *there's a zebra* in epistemically propitious circumstances, circumstances in which it is true and not accidentally true. So although our belief is actually false, if it were true, it would not be accidentally true.

Given the description of the case, though, it is not clear that the relevant belief would not be luckily true if it were true. Whether the belief has that feature depends on various details the description fails to nail down—e.g., the likelihood of Fred's poisoning Zak, and the likelihood of Fred's finding another zebra for the exhibit. By contrast, the case I describe in the text clearly involves a false belief that would not be luckily true were it true.

¹⁴ You might suspect that CR straightforwardly generates a promising pro-Fallibilism argument (in addition to a defense of Fallibilism against certain pro-Infallibilism arguments). The suspicion would be mistaken. For one thing, CR's putative requirement does not also *suffice* for warrant. Suppose you hold the belief, B, that all your beliefs are nonluckily true; B surely lacks warrant, notwithstanding the fact that it meets CR's requirement (cf. Howard-Snyder et al. (2003); they attribute the point to David Vander Laan). To argue from CR to Fallibilism, then, you must add a highly questionable premise—viz., that a false belief could meet the minimally sufficient condition for warrant that results from combining CR's requirement with whatever additional requirements there are on warrant. For another thing, as I'll argue in Sect. 4 below, CR's alleged requirement on warrant seems too strong: it seems that there could be a warranted belief which might have been luckily true had it been true.

¹⁵ For such an argument, see Merricks (1995) and Greene and Balmert (1997).

¹⁶ Cf. Sect. 2 of Howard-Snyder et al. (2003).

¹⁷ Here's a sample "closure" principle for justification: if S (i) justifiedly believes p and (ii) properly infers q from p, then S's belief that q is justified. And here's a sample "closure" principle for knowledge: If S (i) knows p and (ii) properly infers q from p, then S knows q.

¹⁸ For critical discussion of closure principles for justification, see Chap. 6 of Audi (2003). For critical discussion of closure principles for knowledge, see David and Warfield (forthcoming).

2 Merricks's supervenience argument

The most prominent version of the Supervenience Argument is due to Merricks (1995, 1997). Very roughly, **Merricks's Supervenience Argument** goes like this. Suppose there is a warranted false belief. Any such belief could become luckily true without losing its warrant.¹⁹ So, there could be a luckily true warranted belief, a luckily true case of knowledge. But of course there could not be such a case of knowledge. So, there cannot be a warranted false belief: only *true* beliefs can have warrant.

Here is a much more precise reconstruction of Merricks's Supervenience Argument.²⁰ Note first that the argument depends on two plausible assumptions:

Assumption 1: Warrant has a supervenience base—i.e., the facts about whether a given belief has warrant are determined by certain *other* facts about the belief. (Call the properties on which warrant supervenes the **base properties**.²¹)

Assumption 2: If a belief has warrant in one possible world but lacks it in another, **then** the belief exemplifies some base property better in the former than it does in the latter. (In Merricks's terminology, the belief's "overall epistemic situation" [628] is in some respect better in the former than in the latter.)

From Assumptions 1 and 2, the Luck Argument runs as follows:

Premise 1: Suppose, for reductio, that there is a warranted contingently false belief, B.²²

Premise 2: If a belief is warranted and contingently false, **then** it is possible that the belief be luckily true for its subject while exemplifying each base property as well as it actually does.²³

Merricks generalizes to Premise 2 from the following two cases:

[S]uppose for reductio that, in a possible world W, I have the *false warranted belief* that Jones owns a Ford Escort. Now imagine another world W* that differs from W in that—seconds before I form the belief that Jones owns an Escort—Jones' aunt dies, thousands of miles away and unbeknownst to Jones or me, bequeathing Jones an Escort. W* differs from W in all that this difference entails but in no other way... W and W* are alike in the environment—for thousands of miles around—in which I formed my belief, in the reasons for and causes of my belief, in all my psychological states, and so on. There is no defeater for my belief in W* that is not also a defeater for that belief in W. If it will help, we can add that Jones' aunt dies and bequeaths in W as well, a

¹⁹ For simplicity, I assume that a belief can survive a change in its content's truth value. This assumption is in no way crucial to anything I do.

²⁰ See Merricks (1997: 627–631).

²¹ Notably, the Supervenience Argument remains nonpartisan by not taking a stand on *which* properties warrant supervenes on.

²² *Question:* Why restrict Premise 1 to *contingently* false beliefs? *Answer:* Merricks's argument depends on (something like) the thought that any warranted false belief could become luckily true without losing its warrant. But as Merricks explicitly notes (630–631), no *necessarily false* belief could even become *true*, much less become luckily true without losing warrant. Thus, were Premise 1 *not* restricted to contingently false beliefs, the argument would in effect simply assume that no necessarily false belief could have warrant, and so would run the risk of dialectical impropriety (since it would in effect simply assume from the outset a good deal of what it aims to establish).

²³ This is my restatement of the thesis labeled '(1)' in Merricks (1997: 630).

few seconds *after* I formed my belief... I submit that there is *no* difference between the worlds that would constitute a plausible difference in the supervenience base for my belief's warrant. (Merricks 1997: 628–629)

Suppose (for reductio) that I have the false warranted belief, in W_1 , that there is a pear on the table in front of me. I have been deceived by a pear-like hologram, which I mistake for a pear. Consider a nearby possible world, W_2 , differing from W_1 only in that my belief is true and in all this entails... The overall epistemic situation of W_1 , although perhaps different from that of W_2 , is not *better* than that of W_2 . The situations in W_1 and W_2 seem equally poor. Both contain misleading pear holograms. Indeed, if one is better than the other, it would seem to be the one in W_2 , in which, at least, my pear belief is *true*. (Merricks 1997: 629–630)

From Assumption 2 and Premises 1 and 2, we can infer

Lemma: B could have warrant and be luckily true for its subject.

But of course

Premise 3: There could not be a warranted belief which is luckily true for its subject.

What follows, by reductio, is

Restricted Infallibilism: There could not be a warranted *contingently* false belief.

Merricks then augments the foregoing line of thought with the claim that (unrestricted) Infallibilism is more plausible than what he calls

Restricted Fallibilism II: Restricted Infallibilism is true, but there could be a warranted *necessarily* false belief.

Thus, Merricks concludes, we have good reason to accept Infallibilism.

We will begin our evaluation of Merricks's Supervenience Argument by focusing on the judgment that Restricted Fallibilism II is less plausible than (unrestricted) Infallibilism. Let us grant for the moment that Merricks has established Restricted Infallibilism—i.e., the thesis that there cannot be a warranted contingently false belief. I submit that, even under the assumption of Restricted Infallibilism, certain necessarily false beliefs—e.g., a false mathematical belief based on careful yet subtly fallacious reasoning, or some or other of the beliefs involved in the famous Russell paradoxes²⁴—seem at least as likely to *have* warrant as they are to *lack* it. By my lights, then, Merricks's relative plausibility judgment concerning Restricted Fallibilism II and (unrestricted) Infallibilism is wrong: at the relevant dialectical stage, Restricted Fallibilism II seems no less plausible than Infallibilism.

But Merricks's Supervenience Argument fails even if the indicated plausibility judgment is correct. For one thing, as Howard-Snyder et al. (2003) have argued, Premise 2 is no more plausible than its denial.²⁵ To see this, consider a property had by certain of our

²⁴ For a helpful introductory discussion of the Russell paradoxes, see (e.g.) Loux (2006: 30–31). The “property version” of the Russell paradox involves the following two beliefs (each of which is true iff *necessarily* true): (i) there is such a property as *self-exemplification*, and (ii) every property has a complement. Each of these beliefs enjoys a good deal of intuitive plausibility, yet at least one must be false. For if both are true, there is such a property as *non-self-exemplification*. But there can be no such property, since it would exemplify itself iff it did not exemplify itself.

²⁵ What follows in this paragraph is a summary of the objection pressed in Sect. 3.3 of Howard-Snyder et al. (2003).

beliefs: the property of *being such that it would be nonluckily true were it true*. Call this the **Counterfactual Property**. Premise 2 entails that no warranted contingently false belief could have the Counterfactual Property.²⁶ So, Merricks's argument succeeds only if Premise 2 is more plausible than the view that there could be a warranted contingently false belief which has the Counterfactual Property. Unfortunately, the view that there could be such a warranted contingently false belief seems *at least* as plausible as Premise 2. Premise 2 is thus no more plausible than its denial.²⁷

What's worse, Premise 2 seems *false*: some plausible candidates for warranted contingently false belief are such that, in *every* world in which they are luckily true, they exemplify some base properties worse than they in fact do. Consider Sandy, whose brain is—unbeknownst to her—such that she cannot now have tokens of mild pain (though she does grasp the concept *mild pain*). Suppose Sandy comes to believe *falsely* that she is in mild pain—what she feels is in fact an acute tickle.²⁸ Call Sandy's false introspective belief 'B'. Notice that B is only *contingently* false—to make B true, we need only “fix” Sandy's brain so that she can have mild pains, then place her in appropriate circumstances. I submit that Sandy's introspective belief B is a plausible candidate for warranted contingently false belief. At any rate, it is difficult to imagine a *better* candidate for warranted contingently false belief: beliefs like B—viz., introspective beliefs grounded in attentive self-consciousness—are not only typically true, but also typically constitute knowledge (and so, typically have warrant) when true. Supposing that there *are* some plausible candidates for warranted contingently false belief, then, B would seem to be among them. Arguably, though, in *every* world in which B is luckily true, B exemplifies some base property worse than it in fact does. Let me explain.

Any world in which B is luckily true is one in which (i) Sandy's brain has been fixed so that she can have mild pains, *but* (ii) Sandy cannot yet reliably distinguish acute tickles from mild pains. In any such world, B fails to meet the requirement on warrant laid down by the extremely plausible

Safety Principle: S's belief, B, has warrant only if the following is true: if S were to hold a belief as she holds B, S would probably hold a *true* belief in that way—i.e., in most of the nearby possible worlds in which S holds B as she actually does, S holds a *true* belief in that way.²⁹ (For endorsements of claims like the Safety Principle, see, e.g., Pritchard (2005), Sosa (1999), Sainsbury (1997) and Williamson (2000).)

²⁶ Suppose there could be a warranted contingently false belief which has the Counterfactual Property. Then there could be a warranted contingently false belief such that, in *any* world in which the belief is luckily true, it exemplifies some of the base properties worse than it in fact does. That last claim is just the denial of Premise 2.

²⁷ We should note, at least in passing, a relevant observation Merricks makes in his 1997 (footnote 7). According to Merricks, it is plausible that a belief's having a feature like the Counterfactual Property will be due to the belief's having certain other, more fundamental epistemically valuable features. This point seems consistent with the objection just pressed. That objection assumes only that (i) Premise 2 is no more plausible than the view that there could be a warranted false belief which has the Counterfactual Property; and that (ii) a belief's losing the Counterfactual Property entails that the belief does not exemplify some or other base property as well as it once did. By my lights, this assumption is compatible with the view that a belief's having a feature like the Counterfactual Property will be due to the belief's having certain other, more fundamental epistemically valuable properties. (See footnotes 29, 37, and 45 below, where I defend other objections to the Supervenience Argument against Merricks's observation.)

²⁸ I follow others here in assuming a mild “fallibilism” regarding one's beliefs about one's own sensations. For sympathetic discussion of this assumption, see (e.g.) Chaps. 4 and 8 of Williamson (2000), Pojman (2001: 103–104) and Feldman (2003: 55–57).

²⁹ Five points about the Safety Principle. First, proponents of the principle assume it makes sense to talk about “the way” in which (“the basis” on which, “the method” by which) a given belief is held. Granted, there is significant disagreement about the nature of the ways in which (bases on which, methods by which) beliefs are held. (For example, are such ways [bases, methods] individuated only via “internal” properties,

Finally, we may safely assume that B *in fact* meets the requirement laid down by the Safety Principle—that B is *in fact* such that, if Sandy were to hold a belief as she holds B, she would probably hold a *true* belief in that way. (Sandy is capable of having acute tickles, but *not* mild pains; as a result, she is much more likely to hold beliefs involving the concept *acute tickle* than ones involving the concept *mild pain*.) In sum, then, B is a plausible candidate for warranted contingently false belief, *but* is such that in any world in which it is luckily true, it exemplifies some base property—viz., the one involved in the Safety Principle—worse than it in fact does. The foregoing line of thought casts serious doubt on Premise 2, which (recall) says that, for *any* plausible candidate for warranted contingently false belief, there is a world in which the belief is luckily true *yet* exemplifies each base property as well as it actually does.³⁰

3 The improved supervenience argument

To improve on Merricks's Supervenience Argument, we need to replace Premise 2 with something that enjoys more plausibility *and* yields a more direct argument for Infallibilism. Here is a version of the Supervenience Argument whose substitute for Premise 2 has both of these virtues. We start with assumptions similar to those that underwrite Merricks's version:

Assumption 1: Warrant has a supervenience base.

Assumption 2*: A belief loses warrant only if the belief comes to exemplify some base property worse than it once did.

From Assumptions 1 and 2*, the improved version of the Supervenience Argument runs as follows:

Footnote 29 continued

or do “external” properties also figure in such individuation?) All the Safety Principle's proponent needs, though, is the common assumption that there is such a thing as the way in which (the basis on which, the method by which) a given belief is held. (For discussion of some issues relevant here, see, e.g., Pritchard (2005, Sects. 6.2–6.4) and Williamson (2000, Sects. 7.4–7.5).)

Second, the Safety Principle is neutral between Infallibilism and Fallibilism, and so can figure in a dialectically appropriate reply to the Supervenience Argument.

Third, I think most theorists will accept (something like) the Safety Principle; for it seems that the principle will be entailed by any adequate solution to the Gettier Problem—i.e., the challenge of identifying a plausible requirement on warrant that excludes the possibility of “Gettierized” knowledge (cf. Pritchard 2005, Chap. 5).

Fourth, proponents of the Safety Principle assume it makes sense to talk about “most” of the nearby worlds in which a given proposition is true. Because there will (presumably) be infinitely many such worlds, the Safety Principle's proponents will have to assume—with (among others) Van Inwagen (1997)—that some

infinite sets of worlds have a “size,” a measure, that is not determined by their cardinality. Sets of possible worlds are thus analogous to sets of points on a plane, many of which have sizes or measures that are not determined by their cardinalities: one set of points may occupy two square feet, and another set of the same cardinality (the power of the continuum) may occupy one square foot—or a square light-year or a square micron. In my view, it is impossible to make sense of probabilities without making this assumption. (235–236)

Fifth, the objection currently under construction assumes only that (i) a warranted belief will have the feature highlighted by the Safety Principle's consequent, and that (ii) a belief's losing that feature entails that the belief does not exemplify some or other base property as well as it once did. Accordingly, the objection is compatible with the view (see, e.g., Merricks 1997, footnote 7) that a belief's having that feature will be due to its having certain other, more fundamental epistemically valuable properties.

³⁰ It should be clear that this objection to Premise 2 is dialectically appropriate: in no way does it presuppose that Fallibilism is actually true.

Premise 1*: Suppose, for reductio, that there could be a warranted false belief.

Introducing a technical term now will help us avoid verbosity later. Say that a belief **safely** survives a certain change iff the belief survives the change without coming to exemplify some base property worse than it did before (in Merricks’s terminology, without its “overall epistemic situation” worsening in any respect). The argument continues like this:

Premise 2*: If there could be a warranted false belief, **then** there could be a warranted belief that safely becomes luckily true for its subject.

Here is the most promising subargument for Premise 2* I know of:³¹

Suppose Sally’s environment and faculties are *nearly* ideal for gaining perceptual knowledge. The only flaw is an extremely realistic-looking fake apple in the middle of her room. Sally looks at the fake apple just before *t*, and at *t* forms the *false* belief that there is an apple in the room. Call Sally’s false belief ‘*B*’. Suppose Sally still holds *B* at *t**, a time just after *t* at which *B* remains false.

Suppose—for conditional proof—that there could be a warranted false belief. Plausibly, if there could be such a belief, *B* qualifies as one. So, *B* has warrant. Now, suppose there is—at *t*—a small chance that—by *t**—a *real* apple will have rolled quietly into Sally’s room, hidden from her view. Suppose, finally, that the real apple in fact sneaks into Sally’s room by *t**. Plausibly, at *t** in this alternative scenario—a time when *B* is *true*—, *B* exemplifies each base property as well as it *actually* does at *t**—a time when *B* is *false*. But in the alternative scenario, *B* is just luckily true for Sally at *t**. So, Premise 2* is true: **If** there could be a warranted false belief, **then** there could be a warranted belief that safely becomes luckily true for its subject.

From Assumption 2 and Premises 1* and 2*, we can infer

Lemma: There could be a warranted belief which is luckily true for its subject.

But of course

Premise 3: There could not be a warranted belief which is luckily true for its subject.

What follows, by reductio, is

Infallibilism: There could not be a warranted false belief.

This version of the Supervenience Argument—the **Improved Supervenience Argument**—differs from Merricks’s version in at least two important respects: its substitute for Premise 2—Premise 2*—enjoys more plausibility, and yields a more direct argument for Infallibilism, than does Premise 2. These differences constitute improvements on Merricks’s Supervenience Argument.³² But: does the Improved Supervenience Argument succeed?

³¹ The following line of thought is inspired by Merricks’s “Pear Hologram” case (1997: 629–630).

³² Howard-Snyder et al. briefly discuss this argument—or something very similar to it—in Sect. 3.4 of their 2003. Their objection to the argument they consider attacks a subargument for Premise 2* that differs from the one I have offered. As will soon become clear, though, Howard-Snyder et al. (2003) contains raw material for an objection to my favored subargument for Premise 2*. I describe and assess this objection in the next section.

4 Evaluating the improved supervenience argument

I begin by highlighting Premise 2*, along with two claims that figured in my argument for it:

Claim 1: If there could be a warranted false belief, then Sally’s false belief B has warrant.

Claim 2: At t^* in the *alternative* scenario (when B is luckily true), B exemplifies each base property as well as it *actually* does at t^* (when B has warrant).

Premise 2*: If there could be a warranted false belief, **then** there could be a warranted belief that safely becomes luckily true for its subject.

In what follows, I will show that the argument from Claims 1 and 2 to Premise 2* fails, leaving us with no reason whatsoever to believe Premise 2*. Now, while some critics may take issue with Claim 1, I am happy to concede it to the Improved Supervenience Argument’s proponent: if Claim 1 were the argument’s weakest link, I would reckon the argument pretty strong. In what follows, then, I will be focusing on Claim 2.

We should begin by noting that the literature already contains raw material for an initially promising objection to Claim 2. In their (2003), Howard-Snyder introduce and defend what we earlier called

Counterfactual Requirement (CR): A belief has warrant only if the belief is such that it would not be luckily true (i.e., would be nonluckily true) were it true.³³

What we will call **the H-S/F Objection** deploys CR against Claim 2 as follows. Suppose CR is true. Now, recall from the argument for Premise 2* that Sally’s belief, B, is in fact both false *and* warranted—and so, such that it would be nonluckily true were it true—at t^* . Recall also that, at t^* in the alternative scenario, B is luckily true for Sally—and so, is no longer such that it would be nonluckily true were it true.³⁴ At t^* in the alternative scenario, then, B *lacks* warrant, and so does not exemplify each base property as well as it *actually* does at t^* . In sum, if CR is true, then Claim 2 is false. But, finally, CR *is* true. So, Claim 2 is false.

Interestingly, this objection can be parlayed into an objection to Premise 2* itself. Suppose CR is true. Then no warranted belief could safely become luckily true (no luckily true belief is such that it would be nonluckily true were it true). Now, provided that CR is true, it seems that CR would still be true even if there could be warranted false beliefs.³⁵ So, even if there could be warranted false beliefs, no warranted belief could safely become luckily true. But

³³ According to Howard-Snyder et al. (2003: 314), the Counterfactual Requirement “is the minimal link between warrant and truth that precludes Gettier cases; moreover, it does not bring along with it any implausible philosophical theses.” I am about to argue that CR *does* bring some implausible claims in its wake.

³⁴ If B is luckily true, then it might be luckily true were it true, and so is not such that it would be nonluckily true were it true.

³⁵ Let me say something on this claim’s behalf for those who do not find it plausible on its face. Consider the following three propositions:

- If warrant in fact requires adequate evidence, then warrant would still require adequate evidence even if there could be warranted false beliefs.
- If warrant in fact requires grounding in a reliable cognitive faculty, then warrant would still require such grounding even if there could be warranted false beliefs.
- If warrant in fact requires grounding in a properly functioning cognitive faculty, warrant would still require such grounding even if there could be warranted false beliefs.

These three propositions seem correct; and the claim just made in the text seems to stand or fall with them. (Thanks to Nicholas Silins for comments that led me to add this note.)

that last claim conflicts with Premise 2*.³⁶ In sum, CR conflicts with Premise 2*. Because CR is true, we have excellent reason to deny Premise 2*.³⁷

Unfortunately, *neither* of these CR-based objections to the Improved Supervenience Argument succeeds: CR turns out to be dubious, its initial plausibility notwithstanding. In the next paragraph, I will develop an objection to CR. To set the stage, I need to highlight an assumption I will be making about the semantics of counterfactual conditionals.³⁸ I will follow many others in supposing that the truth of a given counterfactual's antecedent and consequent does not suffice for the counterfactual's truth.³⁹ Instead, a given counterfactual is true only if its consequent is true in every *sufficiently close* possible world in which its antecedent is true. On behalf of this view, consider the following cases due to Plantinga (1996: 328–329):

[T]he truth of p and q is not sufficient for the truth of the counterfactual if p then q ... Consider quantum effects: perhaps in fact the photon went through the right slit rather than the left; that is not enough to entail that if it had gone through either slit, it would have gone through the right. I toss the die; it comes up 5. That is not sufficient to entail that if I had tossed the die, it would have come up five. The truth of a counterfactual requires not just that p and $\neg q$ be false in fact; it is also necessary that even if things had been moderately different, it still wouldn't have been the case that p and $\neg q$. To put it in familiar semantical terms, the counterfactual is true only if there is no *sufficiently close* possible world in which p is true but q is not.⁴⁰

Having noted this assumption about the semantics of counterfactuals, I can now present my argument against CR.

According to CR, a belief has warrant only if the belief would not be luckily true (i.e., would be nonluckily true) were it true. CR thus entails that no warranted belief could be luckily true in some world sufficiently close to the actual one. But such warranted beliefs seem possible. Consider the following case, for instance:⁴¹

³⁶ Assuming they are dialectically appropriate, both this objection and Premise 2* are *agnostic* about the question whether there can be warranted false beliefs. In a dialectical context that leaves open a proposition p 's possibility, claims of the form ' $p > q$ ' and ' $p > \text{not-}q$ ' conflict (where ' $p > q$ ' abbreviates 'If p were true then q would be true'). Put differently, unless you are already committed to p 's impossibility, you cannot comfortably accept a claim of the form $[(p > q) \ \& \ (p > \text{not-}q)]$.

³⁷ Notably, the objections described in the last two paragraphs assume only that (i) having the relevant counterfactual property is a plausible condition on warrant; and that (ii) a belief's losing that feature entails that the belief does not exemplify some or other base property as well as it once did. The objection is thus consistent with the view that a belief's having the relevant counterfactual property will supervene on its having certain other, more fundamental epistemically valuable features.

³⁸ Notably, Howard-Snyder et al. seem amenable to the assumption in question (2003: 312–313).

³⁹ See, e.g., Bennett (1974, 2003); Fine (1975); Penczek (1997) and Plantinga (1996, 1997).

⁴⁰ Consider also the following famous case due to Bennett (1974, 378–388):

Suppose that I believe (perhaps on hearsay) that Caspar didn't come to the party and that the party was a bad one, and I say "If Caspar had come, the party would have been a good one." You hear me say this, and you know that Caspar did attend the party and that it was a good party; but you also know that Caspar ruins most parties he attends, and that he *nearly* [my emphasis] ruined this one. It was a good party *despite* the fact that Caspar came to it. Everyone on whom I have tried this example insists that the statement "If Caspar had come, the party would have been a good one" is *not true*; most say that it is *false*; and none will allow that the statement is true but "odd"...

Plausibly, what explains the fact that Bennett's counterfactual is not true is that there are some sufficiently close worlds in which Caspar attends and ruins the party.

⁴¹ The following case is inspired by one Hawthorne describes in his (2004: 4–5).

The standard, “nondeterministic” interpretation of quantum mechanics is correct. Accordingly, there is at noon a nonzero—albeit *infinitesimally small*—objective probability that the particles composing the cup on your desk will promptly behave in such a bizarre way that they no longer compose a cup. At noon, you look at the cup on your desk, and come to believe that there will still be a cup in your office in a few seconds. Naturally enough, your belief is true: the cup on your desk will still be there in a few seconds. Finally—and unbeknownst to you—there is another cup in your office, hidden behind some of your books: you put it there two weeks ago, and have since forgotten about it.

Plausibly, you know that there will still be a cup in your office in a few seconds. If so, then your belief that there will still be a cup in your office in a few seconds—call it ‘B’—has warrant. Now, notice that there are some extremely close worlds in which B is just luckily true—e.g., worlds in which the particles composing the cup on your desk promptly behave so bizarrely that they no longer compose a cup, while the forgotten cup remains intact.⁴² So, since B might have been luckily true had it been true, B is not such that it would have been nonluckily true had it been true. CR thus entails that B lacked warrant—and so, that you do not know that there will still be a cup in your office in a few seconds. In sum, CR conflicts with the plausible claim that you know there will still be a cup in your office in a few seconds, and so seems to place too strong a requirement on warrant.

In light of the foregoing considerations, I submit that the H-S/F Objection’s dependence on the dubious CR spells trouble for the objection. I also think, however, that a very close relative of the H-S/F Objection *does* refute Claim 2. Consider

Weakened Counterfactual Requirement (WCR): A belief has warrant only if the following is true: if the belief were true, the belief would *probably* be nonluckily true—i.e., in most nearby possible worlds in which the belief is true, the belief is not just luckily true.⁴³

Note two facts about WCR. First, if WCR is false, then there could be a warranted belief such that if it were true, it would have no more than an even chance of being nonluckily true. Surely, though, any *warranted* belief—any belief having the property that converts true belief into knowledge—will be such that if it were true, it would have something more than an even chance of being nonluckily true. Accordingly, WCR is extremely plausible. Second, unlike CR, WCR is perfectly compatible with the last paragraph’s claim that you knew there would still be a cup in your office in a few seconds; for we may safely assume that your belief would *probably* have been nonluckily true had it been true.

Armed with WCR, let us return to

Claim 2: At t^* in the *alternative* scenario (when B is luckily true), B exemplifies each base property as well as it *actually* does at t^* (when B has warrant).

⁴² *Pointed Question:* Why think some such worlds—call them ‘Only Forgotten Cup Remains (OFCR)’ worlds—are “extremely close” to the actual one? *Answer:* If we can “obtain” a world, W_1 , from a distinct world, W_2 , by making only an *extremely small change* to W_2 , then W_1 is extremely close to W_2 . Now, notice that we can obtain an OFCR world from the actual one by simply replacing the actual outcome of a certain nondeterministic process with another possible outcome of that process. So, since we can obtain an OFCR world from the actual one by making only an extremely small change to the latter, some OFCR worlds are extremely close to the actual world. (For sympathetic discussion of the “size of change” account of proximity among worlds in play here, see, e.g., Sainsbury (1997) and Van Inwagen (1997).)

⁴³ For a reply to worries about talk of “most” of the nearby possible worlds in which a given proposition is true, see footnote 29 above.

If WCR is true, then Claim 2 is false. Suppose WCR is true. *Question:* At t^* in the alternative scenario, is B such that it would *probably* be nonluckily true were it true? *Answer:* No. In the alternative scenario, there was at t only a small chance that Sally's room would contain a real apple by t^* . At t^* , then, B is such that if it were true, it would be luckily true—and thus, is not such that it would probably be nonluckily true were it true.⁴⁴ So, by WCR, B *lacks* warrant at t^* in the alternative scenario, and so does not exemplify each base property as well as it *actually* does at t^* . WCR thus entails that Claim 2 is false. Because WCR is quite plausible, we now have good reason to deny Claim 2, leaving us with no reason to believe Premise 2* of the Improved Supervenience Argument. Upshot: The best available argument for Infallibilism does not provide good support for that view.⁴⁵

Notably, my objection to the argument from Claims 1 and 2 to Premise 2* suggests a general strategy for criticizing subarguments for Premise 2* like mine—i.e., ones that stem from strong candidates for warranted false belief. Here is the strategy: when facing an argument for Premise 2* that stems from a strong candidate for warranted false belief, look for a plausible candidate for warrant base property which the relevant belief actually exemplifies (because warranted) *but* does not exemplify in the alternative scenario. My deployment of WCR is an instance of this general strategy.⁴⁶

The balance of the paper will assess a new attempt to tilt the scales in favor of *Fallibilism*, the view that false beliefs also can have warrant. The next section will lay out an overlooked puzzle about warrant, and use it to generate a new nonpartisan argument for Fallibilism—the Egalitarianism/Preservationism Argument (EPA). Section 6 will offer an evaluation of the EPA.

5 A puzzle about warrant and an argument for fallibilism

5.1 A puzzle about warrant

The EPA stems from an overlooked puzzle about warrant. This puzzle is the fact that three individually popular claims about warrant turn out to be jointly incompatible. We have already met one piece of the puzzle, viz.,

⁴⁴ If B would be luckily true were it true, then B would probably be luckily true were it true. But if B would probably be luckily true were it true, then B is not such that it would probably be *nonluckily* true were it true.

⁴⁵ Interestingly, unlike CR, WCR cannot be used against Premise 2* itself; for Premise 2* is perfectly compatible with WCR. To see this, note that while a luckily true belief could not be such that it would be nonluckily true were it true, such a belief may nevertheless be such that it would *probably* be nonluckily true were it true.

Note also that my WCR-based objection to the Supervenience Argument assumes only that (i) having the relevant counterfactual property is a plausible condition on warrant; and that (ii) a belief's losing that feature entails that the belief does not exemplify some or other base property as well as it once did. The objection is thus consistent with the view that a belief's having the relevant counterfactual property will be due to its having certain other, more fundamental epistemically valuable features.

⁴⁶ You might suspect that this strategy can be further generalized, as follows: when facing the relevant kind of justification for Premise 2*, you need only cite the property *not being luckily true*, since this property is a plausible candidate for warrant base property which is exemplified by any warranted false belief *but* cannot be exemplified by a luckily true belief. Unfortunately, this suspicion seems mistaken. The problem is that *not being luckily true*—which is equivalent to *being false or being nonluckily true*—is not a plausible candidate for warrant base property. Here are two related reasons for thinking this property—call it **the Disjunctive Property**—is not a warrant base property (cf. Merricks 1997: 628–629; Howard-Snyder et al. 2003: 309). First, it is plausible that a property, P, is a warrant base property only if a belief's exemplifying P entails that the belief has some epistemically positive or valuable feature; but the Disjunctive Property does not meet this requirement. Second, it is plausible that a property is a warrant base property only if the property serves to “shed some light” on the nature of warrant; but the Disjunctive Property does not meet this requirement.

Infallibilism: Only *true* beliefs can have warrant.

A large number of prominent epistemologists are committed to Infallibilism. For one thing, numerous popular accounts of knowledge entail Infallibilism (more on this below). Further, as we have seen, many have developed *nonpartisan* arguments for Infallibilism that are now endorsed by at least a fair number of theorists.⁴⁷ Finally, some theorists accept Infallibilism not because they think it established by some correct analysis of knowledge or successful nonpartisan argument, but rather because they think it somehow strongly suggested by reflection on the post-Gettier literature.⁴⁸

Because my claim that a large number of theorists are committed to Infallibilism depends on my claim that numerous popular theories of knowledge entail Infallibilism, I should pause briefly to justify the latter. I will do so by showing how each of four popular approaches to knowledge entails Warrant Infallibilism.

Consider **Causal Theories** first. According to Causal Theories, the property that distinguishes knowledge that *p* from mere true belief that *p* includes the property *being causally related to the fact that p*.⁴⁹ Of course, a belief that *p* is causally related to the fact that *p* only if *p* is true. So, Causal Theories entail that a belief has warrant only if the belief is true.

According to **Defeasibility Theories**, the property that distinguishes knowledge that *p* from mere true belief that *p* includes the property *being such that there is no fact such that one's belief that p would be unjustified were one to believe the fact*.⁵⁰ If one's belief that *p* is such that there is no fact such that one's belief would be unjustified were one to believe the fact, then *p* must be true. For suppose the last claim's antecedent is satisfied, and consider the denial of *p*. If not-*p* is true, then some fact—viz., not-*p*—is such that one's belief that *p* would be unjustified were one to believe the fact. So, not-*p* is false: *p* is true. Defeasibility Theories entail that a belief has warrant only if the belief is true.

According to **Strong Sensitivity Theories**, the property that distinguishes a case of knowledge from a merely true belief includes the property *being such that it would not be held were it false*.⁵¹ Of course, if a given belief is in fact false, then it is not such that it would not be held were it false. So, Strong Sensitivity Theories entail that a belief has warrant only if the belief is true.

Finally, according to **Strong Safety Theories**, the property that distinguishes an instance of knowledge from a merely true belief includes the property *being such that it would be true*

⁴⁷ For endorsement of such arguments that is representative of at least a fair number of theorists, see Pritchard (2005: 151).

⁴⁸ Consider, e.g., the following passage from David (2001):

...[A]n adequate definition of knowledge requires not only that *S* have a justified true belief, but also that *S*'s justification meet a suitable anti-Gettier condition... Let us refer to justification that meets such a condition as *justification*⁺: *Justification*⁺ is whatever turns true belief into knowledge. Now, it may be that none of the anti-Gettier conditions that have been proposed is fully successful. Still, the work that has been done in this area strongly suggests that any successful condition will have the following feature: A belief cannot meet the condition unless the belief is true. (162)

⁴⁹ See, e.g., Goldman (1967), Armstrong (1973), Dretske (1981) and Nelkin (2000).

⁵⁰ See, e.g., Klein (1996). Actually, the best versions of the Defeasibility Theory are more complicated than this characterization suggests. Fortunately, this paragraph's argument applies equally to those more complicated versions of the Defeasibility Theory, *mutatis mutandis*. I here focus on simpler versions for (what else?) simplicity's sake.

⁵¹ See, e.g., Nozick (1981). Notably, the following point applies equally to so-called "Basis (Method) Relative" Sensitivity Theories, which say (roughly) that a belief has warrant only if the basis on (method by) which the belief is held is such that if the belief were false, it would not be held on that basis (by that method).

were it held.⁵² Of course, if a given belief is in fact false, then it is not such that it would be true were it held. So, Strong Safety Theories entail that a belief has warrant only if the belief is true.

In sum, numerous prominent theorists are committed to Infallibilism for a wide variety of reasons, partisan and nonpartisan.

The remaining two pieces of our puzzle can be introduced more quickly. One is **Preservationism**: If (at t) a belief, B , is **inferentially warranted**—i.e., B depends for its warrant on some other beliefs from whose content B 's subject inferred B 's content—, then (at t) all the beliefs on which B depends for its warrant will themselves have warrant.

In a slogan: “Inference can do no more than *preserve* warrant.” Like Infallibilism, Preservationism enjoys wide acceptance among prominent epistemologists; for it is entailed by the orthodox position on the epistemology of inference, according to which inference cannot *create* or *generate* epistemic commodities like justification, warrant, and knowledge, but can only *transfer* or *transmit* such properties from one (group of) belief(s) to another.⁵³

Finally, consider

Egalitarianism: False (as well as true) beliefs can contribute nonredundantly to another belief's constituting inferential knowledge.

Though it is less popular than Infallibilism and Preservationism, Egalitarianism nonetheless enjoys fairly wide acceptance among prominent theorists. It is endorsed in recent work by (among others) John Hawthorne, Stephen Hetherington, Risto Hilpinen, Peter Klein, Trenton Merricks, Peter Unger, Ted Warfield, and Michael Williams.⁵⁴ In the next section, I will describe and discuss cases such theorists have offered on Egalitarianism's behalf.

Now that the three pieces of our puzzle have been introduced, let us consider an argument that shows them to be jointly incompatible. Suppose that Egalitarianism is true. Then there could be an instance of inferential knowledge, B_1 , such that a false belief, B_2 , makes a nonredundant contribution to its status as inferential knowledge. In such a case, B_1 would depend for its warrant—at least in part—on the false belief, B_2 . Now, suppose also that Preservationism and Infallibilism are true. According to Preservationism, each belief B_1 depends on for its warrant must itself have warrant. So, because B_1 depends on B_2 for its warrant, B_2 must itself have warrant. By Infallibilism, then, B_2 is true. By hypothesis, though, B_2 is false. So, B_2 is both true and false, which is absurd. Infallibilism, Preservationism, and Egalitarianism are jointly incompatible.

We face a puzzle. Each of Infallibilism, Preservationism, and Egalitarianism is a frequently endorsed claim about warrant, which suggests that each can withstand a considerable amount of philosophical scrutiny. As we have just seen, however, the three claims are jointly

⁵² See, e.g., Sainsbury (1997), Sosa (1999), Williamson (2000) and Pritchard (2005). Notably, leading versions of Reliabilism (e.g., Goldman 1986), Proper Functionalism (e.g., Plantinga 1997), and the Virtue Theory (e.g., Zagzebski 1996) all qualify as Strong Safety Theories. Note also that the following point applies equally to so-called “Basis (Method) Relative” Strong Safety Theories, which say (roughly) that a belief has warrant only if the basis on (method by) which the belief is held is such that if the belief were held on that basis (by that method), the belief would be true.

⁵³ For statements of the orthodox position on the epistemology of inference, see Chap. 6 of Audi (2003) and Plantinga (1993: 37, 137–138, 178).

⁵⁴ See Hawthorne (2004: 57), Hetherington (1996: 68), Hilpinen (1988: 163–164), Klein (1996: 106), Merricks (2001: 173–175), Unger (1968), Warfield (2005) and Williams (1978).

incompatible. Solving this puzzle requires that we determine which of the claims should be rejected. So, which claim(s) should we reject?

5.2 The egalitarianism/preservationism argument

Enter the EPA. Because Infallibilism, Preservationism, and Egalitarianism are jointly incompatible, you can argue from any two of these claims to the denial of the third. That is exactly how the EPA works. The EPA's first premise is Preservationism, the thesis that inference can do no more than preserve (transmit, transfer) warrant. Moreover, the argument continues, Egalitarianism is true: a false belief could make a nonredundant contribution to another belief's constituting inferential knowledge. Thus, the argument concludes, Infallibilism is incorrect: false beliefs can have warrant. Essentially, the EPA's proponent maintains that *accepting Fallibilism* is the best way to solve our puzzle about warrant.

At least two important questions arise in the wake of this new nonpartisan argument for Fallibilism.⁵⁵ First, if the argument succeeds, which main theories of knowledge are impugned, and which are confirmed? Second, does the argument succeed? I will devote the rest of this subsection to answering the first question, and will offer an answer to the second question in Sect. 6.

Regarding the first question, we can begin by noting that the argument's conclusion entails that we should reject any theory of knowledge according to which only *true* beliefs can have warrant. The EPA thus entails that we should reject the four theories mentioned above in connection with Infallibilism, and more besides—e.g., species of the **Justified True Belief Theory** according to which only *true* beliefs can be “fully justified,” and that only “fully justified” beliefs constitute knowledge.⁵⁶ Theories confirmed by the EPA include the pre-Gettier Justified True Belief account,⁵⁷ certain post-Gettier species of the Justified True Belief account,⁵⁸ **Weak Sensitivity Theories**—which claim (roughly) that a belief is knowledge only if it would *probably* not be held were it false—, and **Weak Safety Theories**—which claim (roughly) that a belief is knowledge only if it would *probably* be true were it held.⁵⁹

6 Evaluating the EPA

We turn now to the question whether the EPA succeeds. We can make progress toward answering this question by considering the available support for *Egalitarianism*, which seems initially to be the EPA's more questionable premise.

The EPA succeeds only if its proponent says something on Egalitarianism's behalf. This is because there is an initial presumption *against* Egalitarianism, in favor of the view that a belief is inferential knowledge only if all the beliefs it depends on for its epistemic status

⁵⁵ Some may wonder just *how* nonpartisan the Best Solution Argument really is. In particular, some may wonder whether a theorist not already committed to Fallibilism could be at all inclined to accept Egalitarianism. Conclusive evidence for an affirmative answer is the fact that some theorists explicitly committed to *Infallibilism*—e.g., Trenton Merricks and Peter Klein—also accept Egalitarianism (more on this below). (Thanks to Peter Van Inwagen for comments that led me to add this note.)

⁵⁶ See, e.g., Chisholm (1982: 45–47) and Sturgeon (1993).

⁵⁷ For an intriguing recent attempt to defend the original Justified True Belief account against Gettier-type examples, see Weatherston (2003).

⁵⁸ For an interesting recent attempt to develop a Fallibilist Justified True Belief account that sidesteps Gettier-type examples, see McGrew and McGrew (1998).

⁵⁹ For discussion of Weak Sensitivity and Weak Safety Theories, see Kvanvig (2004). Notably, various species of Reliabilism, Proper Functionalism, and the Virtue Theory fall under these headings.

are themselves *true* (cf. Warfield 2005). (Indeed, the dominant conception of inferential knowledge has it that a belief is inferential knowledge only if all the beliefs it depends on for its epistemic status are themselves *cases of knowledge* [cf. Audi 2003, Chap. 7].) Here is one rough-and-ready way to see Egalitarianism's initial implausibility. Any *false* belief will thereby have at least one considerable epistemic defect. A case of inferential knowledge, on the other hand, will not have any considerable epistemic defects (otherwise, it would not constitute knowledge). In light of these simple points, Egalitarianism can look pretty implausible: how could a piece of reasoning that involves something with a considerable epistemic defect yield something that has *no* such defects? On its face, then, Egalitarianism seems wrong: it seems that there could not be inferential knowledge that depends for its epistemic status on a false belief. The EPA succeeds only if Egalitarianism can overcome this initial presumption against it.

Fortunately for the EPA, there *is* support available for Egalitarianism: Egalitarianism is supported by intuitive judgments about a variety of cases described in work by theorists mentioned in Sect. 5.1. According to John Hawthorne, for example,

[i]t is arguable that, contrary to recent epistemological tradition, one can allow that knowledge might be based on false beliefs, so long as this does not make for relevant mistakes—as when a child knows there will be toys under the Christmas tree on the basis of the false belief that Santa Claus is coming or when I know that someone is short on the basis of a perceptually based belief that the person is under four feet tall (when the person is in fact four foot one). (2004: 57)

Consider also the following case due to Peter Klein (call it **Reliable Secretary's Testimony**):

...[C]ases can be constructed so that a belief is known (hence warranted) even though based on a false proposition, when what makes the proposition false is not, so to speak, fatal. For example, suppose that my belief that I have an appointment at 3:00 p.m. on April 10th is based on my... false belief that my secretary told me on April 6th that I had such an appointment. If my secretary told me that I had such an appointment, but she told me that on April 7th (not the 6th), my belief that I have an appointment at 3:00 p.m. can still be knowledge, even though the belief that supports it is false. (1996: 106)

Finally, consider this case due to Ted Warfield (call it **Mistaken Reliable Reporter**):

CNN breaks in with a live report. The reliable reporter says: 'The President is in Utah, speaking to supporters'. I then reason: 'The President is in Utah; therefore he is not attending today's NATO talks in Brussels'. I know my conclusion but my premise is false: the President is in Nevada—unbeknownst to me, he's speaking at a "border rally" at the border of those two states and the speaking platform on which he is standing is definitely in Nevada. (2005: 408, lightly edited)

Each of these examples seems to involve inferential knowledge. Moreover, in each example, it seems that a *false belief* is making a nonredundant contribution to the epistemic status of the inferential knowledge. Thus, each example seems to involve inferential knowledge which depends for its status as such—at least in part—on a false belief. Granted, in each example, we can identify a *true proposition* which has the following two features:

- the subject is (at least) disposed to believe the proposition,

and

- if the subject's inferential belief were based on a belief in that proposition, the inferential belief would still constitute knowledge (other things equal).⁶⁰

What this suggests is that none of the cases of inferential knowledge just described *essentially* depends for its epistemic status on a false belief. But that point is consistent with the view that the relevant instances of inferential knowledge nevertheless *actually* depend for their epistemic status on—in *fact* owe their epistemic status to—the relevant false beliefs. Consider an analogy. A roof may be nonredundantly supported by—and so, actually depend for its position on (owe its position to)—a particular column, *even though* the roof would not collapse were it to lose support from that column (because, say, the column would be immediately replaced with another were it damaged). Likewise, a case of inferential knowledge may receive a nonredundant epistemic contribution from—and so, actually depend for its epistemic status on (owe its epistemic status to)—a particular belief, *even though* the former would still be knowledge were it to lose epistemic support from the latter. So, even if none of the cases of inferential knowledge just described *essentially* depends for its epistemic status on a false belief, each may nevertheless *actually* depend for its epistemic status on a false belief. Arguably, then, the examples just described constitute some justification for Egalitarianism, the view that a false belief can contribute nonredundantly to another belief's being inferential knowledge.

In light of the last paragraph *and* the fact that Preservationism follows from the orthodox position on the epistemology of inference, I submit that the EPA applies new and nontrivial dialectical pressure to Infallibilism. A question now arises as to how long this new pressure on Infallibilism can be sustained. I believe that, while initially promising, the EPA ultimately fails. In the balance of this section, I will offer two different but related explanations of the EPA's failure. Both explanations will stem from the following example—call it **Gettierized Reliable Reporter**:

CNN breaks in with a live report. The reliable reporter says: 'The President is in Utah, speaking to supporters'. I then reason: 'The President is now in Utah; therefore he is not attending today's NATO talks in Brussels'. I know my conclusion but my premise lacks warrant. Unbeknownst to me, the President is speaking at a "border rally" at the Utah-Nevada border. He is speaking from a large stage that has a "Utah-part" and a "Nevada-part". The President walked into Utah as the reporter was making his assertion about the President's location, which was just before I formed my belief that the President is in Utah. I do not know that the President is in Utah: the President could very easily have been in *Nevada* when I formed my belief that he is in Utah. My true belief that the President is in Utah thus falls short of knowledge, and so lacks warrant.

From one angle, the EPA fails because Preservationism is simply false. For cases like Gettierized Reliable Reporter establish that inference can do more than just *preserve* warrant, that there can be an inferentially warranted belief which depends for its warrant (at least in part) on some beliefs which themselves *lack* warrant. In light of such cases, we can see that Preservationism is not plausible on balance: its *prima facie* plausibility does not win through to *ultima facie* plausibility.

Looking from a slightly different angle, the problem with the EPA is that the only available justification for Egalitarianism *undermines* Preservationism. As I noted at this section's

⁶⁰ Plausible candidates for such a proposition in Reliable Secretary's Testimony include [My secretary recently told me that I have a meeting on April 10th]. Plausible candidates for such a proposition in Mistaken Reliable Reporter include [The President is near Utah].

outset, there is a natural initial presumption against Egalitarianism, one that can be overcome only via description and defense of cases like (e.g.) Mistaken Reliable Reporter. But deploying such examples commits one to judging cases like Gettierized Reliable Reporter to be ones in which inference does more than just *preserve* warrant. For if anything, it is more reasonable to think that Gettierized Reliable Reporter refutes Preservationism than that Mistaken Reliable Reporter justifies Egalitarianism. Other things equal, a belief which does not depend for its epistemic status on a false belief will be *at least* as strong a candidate for knowledge as a belief which does depend on a false belief for its epistemic status. So, the inferential belief in Gettierized Reliable Reporter seems to be *at least* as strong a candidate for knowledge as its counterpart in Mistaken Reliable Reporter: for while the latter arguably depends on a false belief for its epistemic status, the former does not. Using cases like Mistaken Reliable Reporter to justify Egalitarianism thus commits one—at least implicitly—to denying Preservationism. In sum, the EPA defeats itself in a somewhat surprising way: the EPA succeeds only if some support is offered for Egalitarianism; but the only available support for Egalitarianism *undermines* Preservationism!

At worst, then, Preservationism is simply false; at best, the only available justification for Egalitarianism undermines Preservationism, thereby making the EPA self-defeating. Either way, the EPA fails.

7 Conclusion

In Sect. 1, I noted that Infallibilism and Fallibilism seem roughly equal in initial plausibility. Sections 2–4 established that the best nonpartisan argument for Infallibilism fails. And Sects. 5–6 established that the only nonpartisan argument for Fallibilism fails. Assuming there is no successful *partisan* argument for either of these views, we can derive the following important result from our findings: the Infallibilist–Fallibilist playing field is currently *level*. Contrary to what many theorists think, we currently lack good reason to prefer Infallibilism to Fallibilism, or to prefer Fallibilism to Infallibilism. For the moment, then, our policy when it comes to theorizing about warrant—and so, knowledge—should be latitudinarian: given the current dialectical situation, we cannot sensibly count commitment to Fallibilism as a strike against a theory of warrant, and the same goes for commitment to Infallibilism. So far as we can presently see, neither Infallibilism nor Fallibilism should constrain our theorizing about warrant.⁶¹

References

- Armstrong, D. (1973). *Belief, truth and knowledge*. New York: Cambridge UP.
 Audi, R. (2003). *Epistemology*. New York: Routledge.
 Audi, R. (1993). *The structure of justification*. New York: Cambridge UP.

⁶¹ Ancestors of this paper were presented at the University of Rochester Graduate Conference in Epistemology (2004), the Central States Philosophical Association (2004), the Central Division of the American Philosophical Association (2006), and the University of Tennessee (2007). Thanks to the audiences on those occasions, and to my three conference commentators—John Shoemaker, Kamper Floyd, and Nicholas Silins, respectively. Thanks also to Robert Audi, Andrew Bailey, Justin Biddle, Thad Botham, David DiQuattro, Chris Green, Jeff Green, Stephen Grimm, Christopher Hill, Mark Jensen, Matt Kennedy, Andrew Moon, Michael Pace, Alvin Plantinga, Thomas Senor, J.R. Shrader, Donald Smith, Peter van Inwagen, Ted Warfield, and several anonymous referees. I am especially grateful to Daniel Howard-Snyder and Todd Long for copious written comments that led to numerous substantial improvements in key parts of the paper.

- Bennett, J. (1974). Counterfactuals and possible worlds. *Canadian Journal of Philosophy*, 4, 381–402.
- Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford UP.
- Bigelow, J. No Possibility of Error (<http://www.arts.monash.edu/phil/department/bigelow/error.html>). David and Warfield: No update available.
- BonJour, L. (1996). Plantinga on knowledge and proper function. In J. Kvanvig (Ed.), *Warrant in contemporary epistemology* (pp. 47–71). Lanham, MD: Rowman & Littlefield.
- Chisholm, R. (1982). *The foundations of knowing*. Minneapolis: U of Minnesota P.
- David, M. (2001). Truth as the epistemic goal. In M. Steup (Ed.), *Knowledge, truth, and duty* (pp. 151–169). New York: Oxford UP.
- David, M., & Warfield, T. (forthcoming). Knowledge closure and skepticism. In Q. Smith (Ed.), *New essays in epistemology*. New York: Oxford UP.
- Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Feldman, R. (2003). *Epistemology*. Upper Saddle River, NJ: Prentice Hall.
- Fine, K. (1975). Critical notice: counterfactuals. By D. Lewis. *Mind*, 84, 451–458.
- Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23, 121–123.
- Goldman, A. (1967). A causal theory of knowing. *The Journal of Philosophy*, 64, 355–372.
- Goldman, A. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard UP.
- Greene, R., & Balmert, N. A. (1997). Two notions of warrant and Plantinga's solution to the Gettier problem. *Analysis*, 57, 132–139.
- Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Oxford UP.
- Hetherington, S. (1996). *Knowledge puzzles*. Boulder, CO: Westview.
- Hilpinen, R. (1988). Knowledge and conditionals. *Philosophical Perspectives*, 2, 157–182.
- Howard-Snyder, D., Howard-Snyder, F., & Feit, N. (2003). Infallibilism and Gettier's legacy. *Philosophy and Phenomenological Research*, 63, 304–327.
- Huemer, M. (2005). Logical properties of warrant. *Philosophical Studies*, 122, 171–182.
- Klein, P. (1996). Warrant, proper function, reliabilism, and defeasibility. In J. Kvanvig (Ed.), *Warrant in contemporary epistemology*. Lanham, MD: Rowman & Littlefield.
- Kvanvig, J. (2004). Nozickian epistemology and the value of knowledge. *Philosophical Issues*, 14, 201–218.
- Lehrer, K. (1965). Knowledge, truth, and evidence. *Analysis*, 25, 168–175.
- Loux, M. (2006). *Metaphysics*. New York: Routledge.
- McGrew, T., & McGrew, L. (1998). Internalism and the collapse of the Gettier problem. *Journal of Philosophical Research*, 23, 239–256.
- Merricks, T. (1995). Warrant entails truth. *Philosophy and Phenomenological Research*, 55, 841–855.
- Merricks, T. (1997). More on warrant's entailing truth. *Philosophy and Phenomenological Research*, 57, 627–631.
- Merricks, T. (2001). *Objects and persons*. Oxford: Oxford UP.
- Nelkin, D. (2000). The lottery paradox, knowledge, and rationality. *Philosophical Review*, 109, 373–409.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard UP.
- Penczek, A. (1997). Counterfactuals with true components. *Erkenntnis*, 46, 79–85.
- Plantinga, A. (1993). *Warrant and proper function*. New York: Oxford UP.
- Plantinga, A. (1996). Respondeo. In J. Kvanvig (Ed.), *Warrant in contemporary epistemology*. Lanham, MD: Rowman and Littlefield.
- Plantinga, A. (1997). Warrant and accidentally true belief. *Analysis*, 57, 140–145.
- Pojman, L. (2001). *What can we know?: An introduction to the theory of knowledge*. Belmont, CA: Wadsworth.
- Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford UP.
- Ryan, S. (1996). Does warrant entail truth? *Philosophy and Phenomenological Research*, 56, 183–192.
- Sainsbury, R. M. (1997). Easy possibilities. *Philosophy and Phenomenological Research*, 57, 907–919.
- Sosa, E. (1999). How to defeat opposition to moore. *Philosophical Perspectives*, 13, 141–153.
- Sturgeon, S. (1993). The Gettier problem. *Analysis*, 53, 156–164.
- Unger, P. (1968). An analysis of factual knowledge. *Journal of Philosophy*, 65, 157–170.
- Van Inwagen, P. (1997). Against middle knowledge. *Midwest Studies in Philosophy*, 21, 225–236.
- Warfield, T. (2005). Knowledge from falsehood. *Philosophical Perspectives*, 19, 405–416.
- Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies*, 115, 1–31.
- Williams, M. (1978). Inference, justification and the analysis of knowledge. *The Journal of Philosophy*, 75, 249–263.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford UP.
- Zagzebski, L. (1994). The inescapability of Gettier problems. *Philosophical Quarterly*, 44, 65–73.

- Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of Knowledge*. New York: Cambridge UP.
- Zagzebski, L. (1999). What is knowledge? In J. Greco & E. Sosa (Eds.), *The Blackwell guide to epistemology*. Malden, MA: Blackwell.