Higher-Order Defeat and the Impossibility of Self-Misleading Evidence

Mattias Skipper

Abstract: *Evidentialism* is the thesis, roughly, that one's beliefs should fit one's evidence. The *enkratic principle* is the thesis, roughly, that one's beliefs should "line up" with one's beliefs about which beliefs one ought to have. While both theses have seemed attractive to many philosophers, they jointly entail the controversial thesis that *self-misleading evidence* is impossible. That is to say, if evidentialism and the enkratic principle are both true, one's evidence cannot support certain false beliefs about which beliefs one's evidence supports. Recently, a number of epistemologists have challenged the thesis that self-misleading evidence is impossible on the grounds that misleading higher-order evidence does not have the kind of strong and systematic defeating force that would be needed to rule out the possibility of such self-misleading evidence. Here I respond to this challenge by proposing an account of higher-order defeat that *does*, indeed, render self-misleading evidence impossible. Central to the proposal is the idea that higher-order evidence acquires its normative force by influencing which *conditional beliefs* it is rational to have. What emerges, I argue, is an independently plausible view of higher-order evidence that has the additional benefit of allowing us to reconcile evidentialism with the enkratic principle.

1. Introduction

Evidentialism is the thesis, roughly, that one's beliefs should fit one's evidence. The *enkratic principle* is the thesis, roughly, that one should see to it that one's beliefs cohere with one's beliefs about which beliefs one *ought* to have. Both theses have seemed attractive to many philosophers. However, they jointly imply a controversial conclusion, namely that a certain kind of *self-misleading evidence* is impossible. That is to say, if evidentialism and the enkratic principle are both true, it is impossible for one's evidence to support certain false beliefs about what one's evidence supports.

Recently, a number of authors have argued that self-misleading evidence is possible on the grounds that *misleading higher-order evidence* does not have the kind of strong and systematic defeating force that would be needed to rule out the possibility of such self-misleading evidence. If they are right, we are left with a seemingly unattractive choice between sacrificing the enkratic principle on the altar of evidentialism, or *vice versa*. In other words, anyone who wants to save both evidentialism and the enkratic principle faces a challenge of explaining why cases of misleading higher-order evidence are, in fact, *not* cases of self-misleading evidence.

The aim of this chapter is to propose a view of higher-order evidence that does indeed render self-misleading evidence impossible. Central to the view is the idea that higher-order evidence acquires its normative significance by influencing which *conditional beliefs* it is rational to have. I will say more to clarify and motivate this idea as we proceed. But what I hope will emerge is an independently plausible view of higher-order evidence that has the additional benefit of allowing us to retain both evidentialism and the enkratic principle.

The rest of the chapter is structured as follows. In section 2, I offer some more precise formulations of evidentialism and the enkratic principle and explain why they jointly imply that a certain sort of self-misleading evidence is impossible. In section 3, I then review what I take to be the most serious challenge, due to Worsnip (2018), against the view that self-misleading evidence is impossible, and I argue that the challenge is unconvincing as it stands. In section 4, I go on to propose a novel view of higher-order evidence, which not only offers a diagnosis of where Worsnip's challenge goes wrong, but also lends positive support to the view that self-misleading evidence is impossible. In section 5, I provide some further motivation of the view by showing how it can help us to make sense of a number of peculiar features of higher-order evidence that are otherwise difficult to understand. Section 6 is a brief summary.

2. Evidentialism, Enkrasia, and Self-Misleading Evidence

Let me begin with a few remarks on terminology and notation. As usual, I will write 'Bp' to say that p is believed by the agent in question. Furthermore, if a is a doxastic attitude towards some proposition, I will use the following shorthands to talk about epistemic rationality and evidential support:

ra: a is rationally permitted	ea: a is sufficiently supported by the evidence
Ra: a is rationally required	Ea: a is decisively supported by the evidence

If a is not rationally *permitted*, I will also say that a is rationally *forbidden*. The operators R and E will be treated as duals of r and e in the usual way: that is, a is rationally required if, and only if, $\sim a$ is not rationally permitted, and a is decisively supported by the evidence if, and only if, $\sim a$ is not sufficiently supported by the evidence. I will not rely on any particular semantics of these four operators, but formally inclined readers may think of them as pairs of possibility and necessity operators from standard epistemic and deontic logic (cf. Hintikka 1962; 1971).

On my usage of the term "evidential support," it is *doxastic attitudes* towards propositions rather than propositions themselves that are said to be (or not to be) supported by the evidence. For ease of exposition, I will also sometimes say about a proposition that it is supported by the evidence, but this should be understood to mean that the relevant *belief* in that proposition is supported by the evidence.

For present purposes, I will be individuating doxastic attitudes in a coarse-grained manner. That is, I will be talking about *belief*, *disbelief*, and *suspension of judgment*, but I will not be talking about *degrees of belief* or *credences*. The reason is that the enkratic principle is traditionally understood in terms binary attitudes rather than graded ones. Of course, graded attitudes like credences can be akratic as well. But to keep matters relatively simple, I will focus on the binary case here.

Consider, then, the following statement of the evidentialist thesis:

Evidentialism: Necessarily,

(i)   *ra ↔ ea*; and

(ii) *Ra ↔ Ea*.

According to this thesis, a doxastic attitude is rationally permitted if, and only if, it is sufficiently supported by the evidence; and it is rationally required if, and only if, it is decisively supported by the evidence.

Three comments about Evidentialism are in order. First, when I say that "the evidence" supports this-or-that doxastic attitude, I always have the *total* evidence in mind. Otherwise, Evidentialism would clearly be false, since different parts of a body of evidence might pull in different directions with respect to the same proposition. Second, note that Evidentialism is a "substantive" requirement of rationality in the sense that it constrains which *individual* doxastic attitudes it can be rational to have. By comparison, the requirement to avoid contradictory beliefs is a "structural" requirement of rationality, because it constrains which *combinations* of doxastic attitudes it can be rational to have. Finally, the view I have labelled "Evidentialism" is surely not the *only* evidentialist thesis one might care about. For example, some authors hold that even the most decisive evidence only ever makes it rationally *permissible* (not required) to adopt this-or-that doxastic attitude towards a given proposition.[1] If they are right, we should only accept the first clause in Evidentialism. However, since both clauses are needed to establish the conclusion that self-misleading evidence is impossible, I will focus on this relatively strong version of Evidentialism.

Is Evidentialism a plausible thesis? On the face of it, it seems clear that a person's evidence somehow plays a role in determining which doxastic attitudes it is rational for the person to

---

[1] See, e.g., Conee and Feldman (1985).

adopt. If we are asked to judge whether someone's beliefs on some matter are rational, it seems perfectly appropriate to consider whether the person has responded in a reasonable manner to the available evidence. But despite its *prima facie* appeal, Evidentialism is subject to ongoing debate. For present purposes, I do not want to enter a detailed discussion of the connection between evidential support and epistemic rationality. I will simply grant Evidentialism to let the challenge, to which I aim to respond, arise.

Consider, next, the following statement of the enkratic principle:

Enkratic Principle: Necessarily,

(i) $R(BRa \rightarrow a)$; and

(ii) $R(BR\sim a \rightarrow \sim a)$.

This thesis roughly says that one's doxastic attitudes should "line up" with one's beliefs about which doxastic attitudes one *ought* to have. More precisely, according to the first clause, one is never permitted to believe that one is required to have a doxastic attitude that one does not have. For example, I am never permitted to believe that "I should believe that it's raining" while failing to believe that "it's raining." Conversely, according to the second clause, one is never permitted to have a doxastic attitude that one believes one is not permitted to have. For example, I am never permitted to believe that "I shouldn't believe that it's raining" while believing that "it's raining."

In contrast to Evidentialism, the Enkratic Principle is a *structural* requirement of rationality, because it constrains which *combinations* of doxastic attitudes it can be rational to have, but says nothing about which *particular* doxastic attitudes it can be rational to have. All the Enkratic Principle demands is that agents maintain a certain kind of coherence between their doxastic attitudes and their beliefs about which doxastic attitudes they ought to have.

Is the Enkratic Principle a plausible thesis? At first blush, it might seem self-evidently true. Just as it seems manifestly irrational to believe Moorean propositions like "it's raining, but I don't believe that it's raining," so it seems manifestly irrational to believe akratic propositions like "it's raining, but I shouldn't believe that it's raining." Yet, despite its *prima facie* appeal, the Enkratic Principle is subject to ongoing debate.[2] Again, however, I will not enter a general discussion of whether epistemic akrasia can be rational. For present purposes, I will simply take the Enkratic Principle for granted alongside with Evidentialism.

Now, a number of authors have recently observed that Evidentialism conflicts with the Enkratic Principle in cases where an agent's total evidence misleads about what it itself supports.[3] That is to say, Evidentialism and the Enkratic Principle jointly imply that a certain kind of self-misleading evidence is impossible. More precisely, Evidentialism and the Enkratic Principle imply the following thesis (a proof can be found in the Appendix):

Impossibility of Self-Misleading Evidence: Necessarily,

(i) $Ea \rightarrow {\sim}eB{\sim}ea$; and

(ii) ${\sim}ea \rightarrow {\sim}eBEa.$

This thesis says, roughly, that there are certain false beliefs *about* what one's evidence supports that cannot be supported *by* one's evidence. More precisely, according to the first clause, if one's evidence decisively supports a given doxastic attitude, it cannot sufficiently support believing that it does not sufficiently support that attitude. For example, if my evidence decisively supports believing that "it's raining," it cannot sufficiently support the false belief that "my evidence doesn't sufficiently support believing that it's raining." Conversely,

[2] Some recent discussions of epistemic akrasia can be found in Christensen (2016), Coates (2012), Dorst (this volume), Greco (2014), Horowitz (2013), Lasonen-Aarnio (forthcoming), Littlejohn (2015), Skipper (forthcoming), and Titelbaum (2015).

[3] Most notably Worsnip (2015; this volume) and Lasonen-Aarnio (forthcoming; this volume). Dorst (this volume), Titelbaum (2015), and Williamson (2011; 2014) also come close to the same point.

according to the second clause, if one's evidence does not sufficiently support a doxastic attitude, it cannot sufficiently support believing that it decisively supports that attitude. For example, if my evidence does not sufficiently support believing that "it's raining," it cannot sufficiently support the false belief that "my evidence decisively supports believing that it's raining."

Thus, the Impossibility of Self-Misleading Evidence effectively rules out two kinds of self-misleading evidence as impossible: (i) evidence that sufficiently supports believing that it does not sufficiently support a doxastic attitude that it, in fact, decisively supports; and (ii) evidence that sufficiently supports believing that it decisively supports a doxastic attitude that it, in fact, does not sufficiently support.

Let me pause to explain, in informal terms, why Evidentialism and the Enkratic Principle imply the Impossibility of Self-Misleading Evidence. We begin by supposing that an agent's total evidence at once decisively supports a doxastic attitude, *a*, and sufficiently supports believing that it does *not* sufficiently support *a*. This supposition amounts to denying the first clause in the Impossibility of Self-Misleading Evidence. Assuming that Evidentialism is true, it then follows that the agent is at once rationally required to adopt *a* and rationally permitted to believe that *a* is not sufficiently supported by her evidence. Hence, she is at once rationally required to adopt *a* and rationally permitted to believe that she is not rationally permitted to adopt *a*.[4] But this contradicts the second clause in the Enkratic Principle. Thus, if Evidentialism and the second clause in the Enkratic Principle are both true, the first clause in the Impossibility of Self-Misleading Evidence must be true as well. The second clause in the

---

[4] This step of the argument requires an auxiliary assumption to ensure that one is rationally permitted to believe that *a* isn't sufficiently supported by one's evidence only if one is rationally permitted to believe that one isn't rationally permitted to adopt *a*. See the Appendix for details.

Impossibility in Self-Misleading Evidence can be established in a similar way using the first clause in the Enkratic Principle.

While Evidentialism and the Enkratic Principle together rule out *certain* sorts of self-misleading evidence, they do not rule out *all* sorts of self-misleading evidence. More specifically, they are compatible with the following thesis:

Possibility of Weakly Self-Misleading Evidence: Possibly,

(i) *ea* & *eB~ea*; or

(ii) *~ea* & *eBea*.

Here the first clause says that one's evidence can at once sufficiently support a doxastic attitude *a* and sufficiently support believing that it does not sufficiently support *a*. For example, my evidence might sufficiently support believing that "it's raining" while sufficiently supporting the false belief that "my evidence doesn't sufficiently support believing that it's raining." The second clause says that one's evidence can at once not sufficiently support *a* and sufficiently support believing that it does sufficiently support *a*. For example, my evidence might at once not sufficiently support believing that "it's raining" and sufficiently support believing that "my evidence does sufficiently support believing that it's raining."

We can sum up the foregoing observations by saying that Evidentialism and the Enkratic Principle are compatible with *weakly* self-misleading evidence, but incompatible with *radically* self-misleading evidence. Of course, it might seem arbitrary or *ad hoc* to maintain that weakly self-misleading evidence is possible, whereas radically self-misleading evidence is not. But in any case, it is worth being clear about what follows directly from Evidentialism and the Enkratic Principle, and what is not.

The connection between Evidentialism, the Enkratic Principle, and the Impossibility of Self-Misleading Evidence will constitute a kind of "dialectical fixed-point" in what follows: anyone who accepts Evidentialism and the Enkratic Principle must accept the Impossibility of

Self-Misleading Evidence as well; and anyone who rejects the Impossibility of Self-Misleading Evidence must reject either Evidentialism or the Enkratic Principle (or both). Eventually, I will argue that we should go for the first option: there is, indeed, good reason to think that self-misleading evidence is impossible. But first, let us consider why some philosophers have thought otherwise.

3.   Putative Cases of Self-Misleading Evidence

Consider the following two stories:

> **Driver's Bias**: John believes himself to be above-average at driving. His belief is strongly supported by the evidence: he has many years of experience as a driver, and has a better crash record than most of his acquaintances. But in reading today's newspaper, John learns about the well-documented *driver's bias*: the tendency, especially among male subjects, to overestimate their own driving skills.

> **Poor Logic**: Sophie and her classmates are asked to prove, independently of each other, whether a logical formula, *T*, is tautological or not. As it happens, Sophie makes a few reasoning errors and draws the wrong conclusion that *T* is non-tautological. All of Sophie's classmates happen to reach the same wrong conclusion, and Sophie is aware of their agreement. What is more, their otherwise competent logic professor makes an occasional blunder and assures the students that they have reached the right conclusion.

What does John's total evidence support after having learnt about the driver's bias? And what does Sophie's total evidence support after having learnt that her logic professor and classmates unanimously agree with her? In order to answer these questions, let us take a closer look at the evidence in each case.

In Driver's Bias, John starts out with a body of evidence about his own driving history, which strongly indicates that his driving skills are above average. Evidence of sort is often said to be of "first order," because it directly concerns the question of whether John is better than average at driving. So, by assumption, John's first-order evidence strongly suggests that he is better than average at driving. However, John then receives some additional evidence about the driver's bias, which indicates that he has overestimated his own driving skills. Evidence of this sort is often said to be of "higher order," because it only seems indirectly relevant to the question of whether John's driving skills are above average.

(Later, in section 5.4, I will offer a more precise characterization of this intuitive distinction between first-order evidence and higher-order evidence. But for now, it suffices to have a merely case-based understanding of the distinction.)

In Poor Logic, Sophie starts out with a body of first-order evidence consisting of her flawed proof, which does *not* support that T is non-tautological. She then receives a body of higher-order evidence consisting of the unanimous agreement of her classmates and logic professor, which indicates that her proof is indeed correct. That is, while her first-order evidence does not support that T is non-tautological, her higher-order evidence indicates that her first-order evidence *does* support that T is non-tautological.

What Driver's Bias and Poor Logic have in common is that the higher-order evidence in each case misleads about what the first-order evidence supports. In Driver's Bias, the first-order evidence in fact supports that John is better than average at driving, but the higher-order evidence suggests otherwise. In Poor Logic, the first-order evidence in fact does *not* support that T is non-tautological, but the higher-order evidence suggests that it *does*.[5]

---

[5] Similar cases of misleading higher-order evidence can be found in Christensen (2010), Elga (2013), Horowitz and Sliwa (2015), Schoenfield (2015; 2016), and several chapters in this volume.

How strong are these seemingly opposing evidential relations in Driver's Bias and Poor Logic? Given the present dialectical context, we are looking to create the most hospitable environment for counterexamples to the Impossibility of Self-Misleading Evidence to arise. Let us therefore suppose that John's first-order evidence is strong enough to decisively support the belief that he is better than average at driving, and let us suppose that his higher-order evidence is at least strong enough to sufficiently support believing that the first-order evidence does not sufficiently support believing that he is better than average at driving. Similarly, let us suppose that Sophie's first-order evidence does not sufficiently support her belief that T is non-tautological, and that the higher-order evidence sufficiently supports believing that the first-order evidence decisively supports believing that T is non-tautological.

We can state these four stipulative assumptions about the evidential relations in Driver's Bias and Poor Logic in a more convenient way by introducing the following shorthands for the propositions that feature in the two cases (where "F" stands for *first-order*, and "H" stands for *higher-order*):

P_F: John is better than average at driving.

P_H: John's first-order evidence sufficiently supports BP_F.

Q_F: The logical formula T is not tautological.

Q_H: Sophie's first-order evidence decisively supports BQ_F.

Given this, the four stipulations become: (i) John's first-order evidence decisively supports believing P_F; (ii) John's higher-order evidence sufficiently supports believing $\sim P_H$; (iii) Sophie's first-order evidence does not sufficiently support believing Q_F; and (iv) Sophie's higher-order evidence sufficiently supports believing Q_H.

Let us now return to the question of what the total evidence in each case supports. Given the above stipulations, one might be tempted to reason as follows:

Naïve argument

Driver's Bias: (1) John's first-order evidence decisively supports BP_F.

          (2) John's higher-order evidence sufficiently supports $B{\sim}P_H$.

          (3) So, John's total evidence decisively supports BP_F and sufficiently supports $B{\sim}P_H$.

Poor Logic: (1) Sophie's first-order evidence does not sufficiently support BQ_F.

          (2) Sophie's higher-order evidence sufficiently supports BQ_H.

          (3) So, Sophie's total evidence does not sufficiently support BQ_F and sufficiently supports BQ_H.

Each conclusion, if true, constitutes a counterexample to the Impossibility of Self-Misleading Evidence: the first conclusion violates the first clause in the Impossibility of Self-Misleading Evidence in virtue of saying that John's total evidence at once decisively supports believing P_F and sufficiently supports believing that it does not sufficiently support believing P_F; and the second conclusion violates the second clause in the Impossibility of Self-Misleading Evidence in virtue of saying that Sophie's total evidence at once does not sufficiently support believing Q_F and sufficiently supports believing that it decisively supports believing Q_F. Thus, if the Naïve Argument is sound, we are forced to give up either Evidentialism or the Enkratic Principle.

However, the Naïve Argument ignores the fact that evidential relations are not in general monotonic (hence the pejorative label "naïve"). That is to say, a body of evidence need not support a doxastic attitude just because a subset of the evidence does. There are many examples of this in the literature of epistemic defeat, and I shall not extend the list here.[6] The initial observation I want to make at this point is just that it would be too hasty to conclude

---

[6] For overviews of different issues related to epistemic defeat, see Kelly (2016) and Koons (2017).

that Driver's Bias and Poor Logic are cases of self-misleading evidence on the basis of considerations about what the first-order evidence and higher-order evidence supports when taken *separately*. What matters is what the first-order evidence and higher-order evidence supports when taken *together*. This much should be uncontroversial.

In a recent paper, Worsnip (2018) offers what may be seen as a refined version of the naïve argument, which purports to show that, even if we take into account potential defeat relations in cases like Driver's Bias and Poor Logic, they still constitute genuine counterexamples to the Impossibility of Self-Misleading Evidence (at least on *some* ways of filling in the details of such cases).

Here is a simple reconstruction of Worsnip's main argument, as it applies to Driver's Bias:[7]

Asymmetry argument

W_1 John's first-order evidence supports P_F more strongly than his higher-order evidence supports $\sim P_F$.

W_2 John's higher-order evidence supports $\sim P_H$ more strongly than his first-order evidence supports P_H.

W_3 So, John's total evidence supports both P_F and $\sim P_H$.

Two preliminary remarks about this argument are in order. First, much of Worsnip's own exposition is devoted to establishing two asymmetry claims that are distinct from W_1 and W_2: first, the claim that the higher-order evidence bears more strongly on P_H than on P_F; and, second, the claim that the first-order evidence (if anything) bears more strongly on P_F than on

---

[7] Worsnip's own discussion centers on a case that is structurally similar to Poor Logic; one where the first-order evidence by itself does not support the relevant first-order proposition (see Worsnip 2018, §4.b). However, his argument applies *mutatis mutandis* to Driver's Bias and other similar cases of misleading higher-order evidence. As far as I can see, nothing of importance hinges on which åarticular case of misleading higher-order evidence is used as a backdrop for the present discussion.

P_H. But note that these two asymmetry claims have no straightforward bearing on the conclusion, W_3, that Worsnip ultimately wants to establish. Thus, I take it that Worsnip wants (or, in any case, *needs*) to establish two asymmetry claims akin to W_1 and W_2 that *do* bear on W_3.

Second, note that W_3 is not a clear-cut counterexample to the Impossibility of Self-Misleading Evidence, since it says nothing about how strongly John's total evidence supports P_F and $\sim P_H$ respectively. However, I do not want to resist the Asymmetry argument on the grounds that the total evidence in cases like Driver's Bias and Poor Logic is ever only *weakly* self-misleading. So, for present purposes, I will simply grant that W_3, if true, indeed constitutes a genuine counterexample to the Impossibility of Self-Misleading Evidence. For the same reason, I will also simply concede the two asymmetry claims W_1 and W_2.

Even so, I find the Asymmetry Argument unconvincing as it stands. The problem is that W_1 and W_2 are claims about how the first-order evidence and higher-order evidence bears on P_F and P_H when taken *separately*, whereas W_3 is a claim about how the first-order evidence and higher-order evidence bears on P_F and P_H when taken *in conjunction*. This raises much the same worry that led us to reject the Naïve Argument in the first place. Just as the Naïve Argument draws a conclusion about what the total evidence supports from premises about what the first-order evidence and higher-order evidence supports when taken separately, so the Asymmetry Argument draws a conclusion about what the total evidence supports from premises about what the first-order evidence and higher-order evidence supports when taken separately.

Is there a way to modify the Asymmetry Argument so as to avoid this problem? One could introduce certain additional premises that would ensure that the presence of the first-order evidence does not undermine the bearing of the higher-order evidence on P_F and P_H, and that the presence of the higher-order evidence does not undermine the bearing of the first-

order evidence on P_F and P_H. That is, one could try to amend the Asymmetry Argument with the following premises:

No Bottom-Up Undercutting: The first-order evidence does not undercut the support relation between the higher-order evidence and $\sim P_H$.

No Top-Down Undercutting: The higher-order evidence does not undercut the support relation between the first-order evidence and P_F.

Here the term "undercutting" is used in the standard way (cf. Pollock 1974): roughly, if a body of evidence supports a given doxastic attitude, an undercutting defeater is any body of evidence in light of which the original body of evidence no longer supports that attitude. Thus, No Bottom-Up Undercutting amounts to the claim that the higher-order evidence continues to speak against P_H in the presence of the first-order evidence, and No Top-Down Undercutting amounts to the claim that the first-order evidence continues to speak in favor of P_F in the presence of the higher-order evidence. By combining W_1 and W_2 with No Bottom-Up Undercutting and No Top-Down Undercutting, we have what looks like a strong argument for W_3.

However, there is strong independent reason to reject No Top-Down Undercutting; or so I will argue. Of course, even if No Top-Down Undercutting is false, one might still try to *weaken* No Top-Down Undercutting in various ways by allowing for undercutting defeat to occur as long as the defeat is sufficiently weak (so as not to undermine W_3). But the considerations I will offer against No Top-Down Undercutting are going to speak equally against such weakened versions of that principle. Hence, if the considerations below are correct, the strategy of weakening No Top-Down Undercutting to save W_3 is not going to succeed.

4. Higher-Order Defeat and the Impossibility of Self-Misleading Evidence

So far, I have deliberately avoided all talk about what the total evidence in Driver's Bias and Poor Logic *seems* to support from an intuitive or pre-theoretical standpoint. The reason is dialectical: I think Worsnip's challenge to the Impossibility of Self-Misleading Evidence can be resisted without appealing to such intuitive judgments.

Yet, I take it to be a widely shared intuition that John's total evidence in fact does not sufficiently support believing that his driving skills are above average. Likewise, I take it to be intuitive to say that Sophie's total evidence *does* sufficiently support believing that T is non-tautological. Thus, for what they are worth, these intuitive verdicts suggest that Driver's Bias and Poor Logic are not cases of self-misleading evidence. More importantly, however, I also think we can make good *theoretical* sense of these intuitions. Indeed, as I will argue below, there is an independently plausible view of higher-order evidence available that not only vindicates our intuitive verdicts in cases like Driver's Bias and Poor Logic, but also allows us to maintain that self-misleading evidence is impossible.

As a first step, I want to review what I take to be a compelling diagnosis, due to Christensen (2010; 2011), of what gives rise to our intuitive judgments in cases like Driver's Bias and Poor Logic. Christensen's central observation is that agents who possess strong misleading higher-order evidence seem forced into a kind of *dogmatic* or *question-begging* reasoning, if they maintain their original beliefs in light of the higher-order evidence. Take John as an example: if he, after having received the higher-order evidence, continues to maintain that his first-order evidence supports P_F, it must be because he takes the higher-order evidence to be misleading. After all, if the higher-order evidence had *not* been misleading, the first-order evidence would not have supported P_F. Yet, in assuming that the higher-order evidence is misleading, John seems to beg the question in much the same way as someone who disregards a body of evidence on the grounds that it opposes his or her prior opinions. Thus,

to avoid this sort of dogmatic or question-begging reasoning, John cannot continue to maintain that his first-order evidence supports P_F. Much the same goes for Sophie: just as John cannot reasonably *continue* to maintain that his first-order evidence supports P_F, she can reasonably *begin* to maintain that her first-order evidence supports Q_F in light of the higher-order evidence.

We can unpack this diagnosis a little further by asking how likely John and Sophie should consider P_F and Q_F to be on their respective bodies of first-order evidence *before* versus *after* having received their respective bodies of higher-order evidence. By assumption, John should consider P_F quite likely on the first-order evidence *before* having received the higher-order evidence. However, the foregoing considerations suggest that he should not consider P_F to be likely on the first-order evidence *after* having received the higher-order evidence. If he did, he would disregard the higher-order evidence in what looks like a dogmatic manner. Much the same goes for Sophie: she should not consider Q_F to be likely on the first-order evidence *before* receiving the higher-order evidence, but she may reasonably consider Q_F to be likely on the first-order evidence *after* having received the higher-order evidence.

This seems to suggest that No Top-Down Undercutting must be rejected. Recall that, according to No Top-Down Undercutting, the evidential relation between John's first-order evidence and P_F cannot be undermined by his higher-order evidence, and the evidential relation between Sophie's first-order evidence and Q_F can likewise not be undermined by the presence of her higher-order evidence. Yet, this is precisely what we seem to have denied on the grounds that John and Sophie would otherwise end up assimilating their respective bodies of higher-order evidence in an unreasonable manner. This also means that we have a diagnosis of where the Asymmetry Argument goes wrong: as explained in the previous section, the inference from W_1 and W_2 to W_3 presupposes No Top-Down Undercutting and No Bottom-

Up Undercutting. So, if No Top-Down Undercutting is false, the Asymmetry argument does not go through.

However, the rejection of No Top-Down Undercutting not only allows us to resist the Asymmetry Argument, but also lends more direct support to the view that self-misleading evidence is impossible. To see why, let us reformulate the previous claims about the normative impact of John and Sophie's respective bodies of higher-order evidence in terms of the changes that it should make to their *conditional doxastic attitudes* towards P_F and Q_F.

We can think of an agent's conditional doxastic attitudes as reflecting the way in which the agent takes different bodies of evidence to bear on different propositions. For example, my doxastic attitude towards the proposition "it has recently been raining" conditional on "the streets are wet" is *belief*, since I take the fact that the streets are wet to be strong evidence that it has recently been raining. By contrast, my doxastic attitude towards "it has recently been raining" conditional on "the streets are dry" is *disbelief*, since I take the fact that the streets are dry to be strong evidence against recent rain. Finally, my doxastic attitude towards "it has recently been raining" conditional on "Paris is the capital of France" is *suspension of judgment*, since I take the fact that Paris is the capital of France to have no significant bearing on whether or not it has recently been raining.[8]

The notion of a conditional (all-or-nothing) belief plays much the same role in the present context as the notion of *conditional credence* plays in a Bayesian context. Just as conditional credences reflect the way in which Bayesian agents take different bodies of

_____

[8] Of course, these conditional beliefs will be contingent on the possessed background information, just as an agent's conditional credences are contingent on background information in a Bayesian framework.

evidence to bear on different propositions, so conditional beliefs reflect the way in which "non-Bayesian" agents take different bodies of evidence to bear on different propositions.[9]

We can then ask: how, if at all, should John and Sophie revise their conditional doxastic attitudes towards P_F and Q_F in light of their respective bodies of higher-order evidence? We have said that, before John receives the higher-order evidence, his doxastic attitude towards P_F conditional on the first-order evidence is belief. By assumption, this conditional belief is rational, since his first-order evidence is assumed to support P_F. However, after having received the higher-order evidence, it is no longer rational for him to hold this conditional belief towards P_F, since the higher-order evidence indicates that his first-order evidence does not support P_F. Thus, there is a rational pressure for John to revise his conditional doxastic attitude towards P_F from belief to suspension of judgment (or perhaps even disbelief, depending on how we fill in the details of the case).

A similar story can be told about Sophie: before she receives the higher-order evidence, her doxastic attitude towards Q_F conditional on the first-order evidence is belief. By assumption, this conditional belief is irrational, since her first-order evidence is assumed not to support Q_F. However, after having received the higher-order evidence, it *becomes* rational for Sophie to hold this conditional belief towards Q_F, because the higher-order evidence strongly indicates that her first-order evidence indeed supports Q_F. Thus, while John's higher-order evidence *creates* a rational pressure to revise his conditional belief, Sophie's higher-order evidence *alleviates* an existing rational pressure to give up her conditional belief.

As flagged in the introduction, the lesson I want to draw from these considerations is that *higher-order evidence acquires its normative force by influencing which conditional doxastic*

---

[9] This obviously raises some further questions about how an agent's *conditional* beliefs should relate to the agent's *unconditional* beliefs. However, we need not settle this matter here. For present purposes, it suffices to have an intuitive grasp of the notion of a conditional belief.

attitudes it is rational to have. In Driver's Bias, the higher-order evidence gets to have a normative impact on John's doxastic attitude towards P_F, because it requires him to revise his conditional doxastic attitude towards P_F. In Poor Logic, the higher-order evidence gets to have a normative impact on Sophie's doxastic attitude towards Q_F, because it removes an existing requirement to revise her conditional doxastic attitude towards Q_F.

This also means that we have a way of vindicating our intuitive verdicts in Driver's Bias and Poor Logic. John's total evidence sufficiently (if not decisively) supports believing $\sim P_H$, and does not sufficiently support believing P_F. Sophie's total evidence sufficiently (if not decisively) supports believing Q_H, and sufficiently supports believing Q_F. Neither total body of evidence is self-misleading. Hence, Driver's Bias and Poor Logic do not constitute counterexamples to the Impossibility of Self-Misleading Evidence.

Obviously, even if cases like Driver's Bias and Poor Logic are not examples of self-misleading evidence, there might be *other* such examples. In particular, I have said nothing to sway those philosophers who want to reject the Impossibility of Self-Misleading Evidence on the grounds that one can lack access to what one's evidence *is*.[10] The question of whether such access failures are possible is firmly rooted in the ongoing dispute between internalists and externalists about epistemic rationality, which lies well beyond the scope of this paper. But I hope to have offered some reasons to think that the Impossibility of Self-Misleading Evidence should not be rejected on the grounds that one can receive strong misleading higher-order evidence about what one's first-order evidence supports.

Before we proceed, I should note that the idea that agents sometimes ought to revise their conditional beliefs is not a new one. For instance, Lange (1999) has argued that a number of

---

[10] Authors who have pushed this line of argument include Lasonen-Aarnio (forthcoming) and Worsnip (2018, §4.a), both of whom rely on anti-luminosity considerations, which have originally been used by Williamson (2000, ch. 4; 2015) to defend a broadly externalist stance on epistemic notions like knowledge, evidence, and epistemic justification.

central problems in Bayesian confirmation theory can be resolved if we allow for certain credal changes to come about as the result of revising one's conditional credences (or "confirmation commitments" as he follows Levi (1980) in calling them). Along similar lines, Brössel and Eder (2014) and Rosenkranz and Schulz (2015) have suggested that certain types of peer disagreement should lead the disagreeing parties to revise their credences in the disputed proposition conditional on the shared evidence. The idea, then, that certain types of doxastic changes should come about as the result of revising one's conditional doxastic attitudes has already been put to use in a few different contexts. What I hope to have shown here is that the idea can serve as the basis of a quite general view about the normative significance of higher-order evidence, which allows us to reconcile Evidentialism and the Enkratic Principle.

5. Further Explanatory Attractions

With the core proposal on the table, I now want to provide some further motivation of the proposal by showing how it can help us to understand various peculiar features of higher-order evidence that are otherwise difficult to make sense of. Doing so will also give me an opportunity to clarify and elaborate some aspects of the proposed view.

5.1 The Retrospectivity of Higher-Order Evidence

The first property I want to focus on is the *retrospective* aspect of higher-order evidence. As number of authors have pointed out, someone who receives a higher-order defeater thereby acquires a reason to think that his or her doxastic state was irrational even *before* receiving the higher-order defeater.[11] Take again John as an example: when he receives the information

---

[11] See, e.g., Christensen (2010), Lasonen-Aarnio (2014), and DiPaolo (2016).

about the driver's bias, he thereby gets a reason to think that his belief was never supported by his first-order evidence *in the first place*.

By contrast, first-order evidence does not display this sort of retrospectivity. If, for example, I believe that it is raining outside based on the testimony of a reliable friend, but I then look out the window and see that it does *not*, in fact, rain, I do not thereby get a reason to think that it was irrational of me to believe as I did before looking out the window. Likewise, if I believe that the wall in front of me is red based on its reddish appearance, but I then learn that the wall is merely lit up by a red spotlight, I do not thereby get a reason to think that it was irrational of me to believe as I did before learning about the red spotlight.

Why do higher-order defeaters display this kind of retrospectivity, when first-order defeaters do not? The proposed view offers a simple explanation: the reason why higher-order defeaters have a retrospective aspect is that they work by indicating that one's original conditional doxastic attitude was irrational all along. And the reason why first-order defeaters do *not* have a retrospective aspect is that they do *not* work by indicating that one's original conditional doxastic attitude was irrational to begin with. On the present picture, this is why higher-order evidence has a distinctive retrospective aspect.

5.2 The Agent-Relativity of Higher-Order Evidence

The second peculiar feature of higher-order evidence is that its normative significance in many cases seems to depend on *who possesses it*. To illustrate this kind of agent-relativity, consider a case adapted from Christensen (2010, p. 187):

> **Arithmetic on Drugs**: You and I decide to calculate the square root of 121 independently of each other. Unbeknownst to us, we both settle on the answer "11" and hence form the belief that "$\sqrt{121} = 11$." However, upon having performed our respective calculations, we

both learn that I have been given a reason-distorting drug that subtly, but significantly, impairs my ability to perform even simple arithmetic calculations.

Here it seems that the higher-order evidence (that is, the information that I have been given a reason-distorting drug) has very different normative implications for you and me: while I should lose confidence in my belief, there is no apparent reason for you to do the same. After all, the fact that someone else than yourself is drugged seems utterly irrelevant to the question of whether the square root of 121 equals 11.

As Christensen (2010) and Kelly (2005) have pointed out, this sort of agent-relativity is at least initially a bit puzzling. We are used to thinking that the question of how an agent should respond to a given body of does not depend for its answer on who the agent is. Of course, the answer might depend on the agent's *background information*. But the kind of agent-relativity at play in Arithmetic on Drugs remains even if we assume that you and I have the same background information. Thus, higher-order evidence seems to give rise to a kind of agent-relativity that differs importantly from the kind of relativity to background information with which we are familiar.

What explains this distinct sort of agent-relativity? Again, the proposed view offers a simple answer: in Arithmetic on Drugs, the difference between you and me is that I get a reason to doubt that my first-order evidence supports that "$\sqrt{121} = 11$," whereas you do *not* get a reason to doubt that your first-order evidence supports that "$\sqrt{121} = 11$." Accordingly, I should revise my conditional belief towards "$\sqrt{121} = 11$," whereas you should not. On the present picture, this is why the same body of higher-order evidence may have very different normative implications for different agents.

5.3 The Insignificance of Hypothetical Higher-Order Evidence

The third property that I want to consider is that the normative significance of a body of higher-order evidence sometimes depends on whether it is regarded as *actually* obtaining or as merely *hypothetically* obtaining. To illustrate this phenomenon, consider another case adapted from Christensen (2010, §4):

> **Cognitive Impairment**: Brenda is a formidable scientist, who is going to carry out an experimental test of a hypothesis, h, next week. The experiment has two possible outcomes, o_1 and o_2. While o_1 would strongly confirm h, o_2 would strongly disconfirm h. Being a formidable scientist, Brenda is well aware of these evidential relations. But she is also aware that she will suffer from a cognitive impairment next week that will make her unable to give an accurate assessment of how the experimental results bear on h.

Following Christensen, let us ask two questions about Brenda. First, how confident should Brenda *now* be in h conditional on o_1 and the fact that she will be cognitively impaired next week? Intuitively, very confident! After all, the fact that she will be cognitively impaired next week seems utterly irrelevant to the question of whether o_1 supports h. Second, if Brenda actually learns o_1 next week, how confident should she *then* be that h is true? Intuitively, *not* very confident! After all, she will then be aware that she is cognitively impaired and thus unable to give an accurate assessment of the experimental results. If we put these two verdicts together, we end up saying that Brenda's doxastic attitude towards h next week should not match her current doxastic attitude towards h conditional on o_1 and the fact that she will be cognitively impaired next week.

This is a striking result. We are used to thinking that an agent's credence in a proposition, p, *after* having acquired a body of evidence, should match the agent's *prior* credence in p conditional on that evidence. In other words, we are used to thinking that the normative

significance of a body of evidence does not depend on whether it is regarded as *actually* obtaining or as merely *hypothetically* obtaining. As Zhao et al. (2012) point out, this idea lies at the heart of orthodox Bayesianism, since the rule of Updating by Conditionalization effectively amounts to the claim that one's credences after having learnt that such-and-such *is* the case should match one's prior credences on the supposition that such-and-such *were* the case.

We can illustrate this familiar Bayesian idea with a standard case of first-order defeat:

Defective Experiment: Joe is a formidable scientist who is going to carry out an experimental test of a hypothesis, *h*, next week. The experiment has two possible outcomes, o_1 and o_2. While o_1 would strongly confirm *h*, o_2 would strongly disconfirm *h*. Being a formidable scientist, Joe is well aware of these evidential relations. However, while he is conducting the experiment, Joe learns that one of the key measurement devices in the experimental setup is defective.

Let us again ask two questions about Joe. First, how confident should Joe *now* be in *h* conditional on o_1 and the fact that one of the key measurement devices will be defective next week? Intuitively, not very confident! After all, the defective instrument would render the entire experiment unreliable. Second, if Joe actually learns that the outcome of the experiment is o_1, and that one of the measurement devices is defective, how confident should he *then* be in *h*? Intuitively, not very confident! It seems, then, that Joe's doxastic attitude towards *h* after having conducted the experiment should indeed match his current doxastic attitude towards *h* conditional on o_1 and the fact that one of the key measurement devices will be defective next week. Hence, the normative significance of the evidence about the defective measurement device does not seem to depend on whether it is learnt or supposed.

What explains this difference between first-order evidence and higher-order evidence? The proposed view offers the following explanation: when Brenda learns that she is drugged, she should not simply respond to this higher-order evidence by conditionalizing on it. She should also revise her conditional doxastic attitude towards h (conditional, that is, on o_1). Naturally, then, her resulting doxastic attitude towards h will not match her prior doxastic attitude towards h conditional on o_1 and the fact that she will be cognitively impaired next week.

By contrast, when Joe learns that one of the key measurement devices is defective, he should simply respond to this first-order evidence by conditionalizing on it, which means that his resulting doxastic attitude towards h *will* match his prior doxastic attitude towards h conditional on o_1 and the fact that one of the key measurement devices is defective next week. On the present picture, this is why higher-order evidence may depend for its normative significance on whether it is regarded as learnt or supposed, whereas first-order evidence does not.

5.4 The Indirectness of Higher-Order Evidence

The final property I want to discuss goes back to our initial characterization of the distinction between first-order evidence and higher-order evidence in section 4. The idea there was that first-order evidence bears *directly* on the proposition at hand in a way that higher-order evidence does not. But what, exactly, does this talk of "direct" and "indirect" evidence amount to?

The proposed view suggests the following answer: higher-order evidence is "indirect" in the sense that it influences which doxastic attitudes it is rational to have *by* influencing which conditional doxastic attitudes it is rational to have (conditional, that is, on the relevant first-order evidence). For example, in John's case, the evidence about the driver's bias bears

indirectly on P_F by requiring him to revise his doxastic attitude towards P_F conditional on the evidence about his driving history.

By contrast, first-order evidence is "direct" in the sense that it influences which doxastic attitudes it is rational to have simply by requiring one to conditionalize on it. For example, the evidence about John's driving history bears directly on P_F by requiring him to revise his doxastic attitude towards P_F as a result of conditionalizing on this evidence. On the present picture, this is what the intuitive distinction between "direct" and "indirect" evidence amounts to.

6. Concluding remarks

I have argued in this paper that higher-order evidence acquires its normative force by influencing which conditional doxastic attitudes it is rational to have. If this is correct, we can maintain that cases of misleading higher-order evidence are not cases of *self-misleading* evidence. This would be an important result, since Evidentialism and the Enkratic Principle jointly imply that self-misleading evidence is impossible. The considerations put forth in this paper thus give us a way of reconciling Evidentialism with the Enkratic Principle. None of this shows that these principles are true. Even if what I have said is basically correct, there might obviously be independent reasons to reject Evidentialism or the Enkratic Principle. But at the very least, I hope to have shown that Evidentialism, the Enkratic Principle, and the proposed view of higher-order evidence form an attractive package.

The present paper is clearly just a first step towards a complete theory of the normative significance of higher-order evidence. In particular, I have yet to give a detailed and precise account of how, exactly, a given body of higher-order evidence should influence an agent's conditional doxastic attitudes. I suspect that the various peculiar features of higher-order

evidence discussed in section 5 may inspire such an account. But the finer details are left for future work.

Appendix

This appendix gives a formal derivation of the Impossibility of Self-Misleading Evidence from Evidentialism and the Enkratic Principle. Let '$\Box$' represent conceptual necessity, and let '$\Diamond$' be its dual. The following auxiliary assumption is needed for the derivation to go through:

rB-**closure:** $\Box(\Box[p \rightarrow q] \rightarrow [rBp \rightarrow rBq])$.

According to rB-closure, one is always rationally permitted to believe the necessary consequences of what one is rationally permitted to believe. This is obviously not an uncontroversial assumption. But given the dialectics of the present paper, I will grant it to allow for the derivation to go through.[12]

Lemma 1. The Enkratic Principle (ii), Evidentialism, and rB-closure jointly entail the Impossibility of Self-Misleading Evidence (i).

Proof.

1. $\Box R(BR{\sim}a \rightarrow {\sim}a)$ (Enkratic Principle (ii))

2. $\Diamond{\sim}(Ea \rightarrow {\sim}eB{\sim}ea)$ (*for reductio*)

3. $\Diamond(Ea \,\&\, eB{\sim}ea)$ (2)

4. $\Diamond(Ra \,\&\, rB{\sim}ra)$ (3, Evidentialism, rB-closure)

---

[12] For further discussion of epistemic closure principles, see Klein (1981; 1995) and Schechter (2013).

5. $\lozenge r(a \mathbin{\&} B{\sim}ra)$ $\hfill(4)$

6. $\lozenge{\sim}R(B{\sim}ra \rightarrow {\sim}a)$ $\hfill(5)$

7. $\square(Ea \rightarrow {\sim}eB{\sim}ea)$ $\hfill(\textit{reductio}\text{ from 1, 2, and 6})$

Lemma 2. Evidentialism, the Enkratic Principle (i), and *rB*-closure jointly entail the Impossibility of Self-Misleading Evidence (ii).

Proof. Similar to that of Lemma 1.

Theorem 1. Evidentialism, the Enkratic Principle, and *rB*-closure jointly entail the Impossibility of Self-Misleading Evidence.

Proof. Immediate from Lemmas 1 and 2.

References

Brössel, P. and A.-M. Eder (2014): "How to Resolve Doxastic Disagreement." In: *Synthese* 191, pp. 2359-81.

Christensen, D. (2010): "Higher-Order Evidence." In: *Philosophy and Phenomenological Research* 81, pp. 185-215.

—— (2011): "Disagreement, Question-Begging, and Epistemic Self-Criticism." In: *Philosophers' Imprint* 11.

—— (2016): "Disagreement, Drugs, Etc.: From Accuracy to Akrasia." In: *Episteme* 13, pp. 397-422.

Coates, A. (2012): "Rational Epistemic Akrasia." In: *American Philosophical Quarterly* 49(2), pp. 113-24.

Conee, E. and R. Feldman (1985): "Evidentialism." In: *Philosophical Studies* 48, pp. 15-34.

—— (2004): *Evidentialism*. Oxford: Oxford University Press.

DiPaolo, J. (2016): "Higher-Order Defeat is Object-Independent". In: *Pacific Philosophical Quarterly*. Early view.

Dorst, K. (this volume): "Higher-Order Uncertainty." In M. Skipper and A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, Oxford University Press.

Elga, A. (2013): "The Puzzle of the Unmarked Clock and the New Rational Reflection Principle." In: *Philosophical Studies* 164(1), pp.127–39.

Greco, D. (2014): "A Puzzle About Epistemic Akrasia." In: *Philosophical Studies* 167, pp. 201-219.

Greco, D. and B. Hedden (2016): "Uniqueness and Metaepistemology." In: *The Journal of Philosophy* 113, pp. 365-95.

Hintikka, J. (1962): *Knowledge and Belief*. Ithaca: Cornell University Press.

—— (1971): "Some Main Problems of Deontic Logic." In: *Deontic Logic: Introductory and Systematic Readings*, R. Hilpinen (ed). Studies in Epistemology, Logic, Methodology, and Philosophy of Science 33, Springer Dordrecht, pp. 59-104.

Horowitz, S. (2013): "Epistemic Akrasia." In: *Noûs* 48, pp. 718-44.

Horowitz, S. and P. Sliwa (2015): "Respecting *all* the Evidence." In: *Philosophical Studies* 172, pp. 2835-58.

Kelly, T. (2005): "The Epistemic Significance of Disagreement." In: *Oxford Studies in Epistemology* 1, J. Hawthorne and T. Gendler (eds.). Oxford: Oxford University Press, pp. 167-96.

—— (2016): "Evidence." In: *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), E. Zalta (ed.).

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

—— (1995): "Skepticism and Closure: Why the Evil Genius Argument Fails." In: *Philosophical Topics* 23, pp. 213–36.

Koons, R. (2017): "Defeasible Reasoning." In: *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), E. Zalta (ed.).

Lange, M. (1999): "Calibration and the Epistemological Role of Bayesian Conditionalization." In: *The Journal of Philosophy* 96, pp. 294-324.

Lasonen-Aarnio, M. (2014): "Higher-Order Evidence and the Limits of Defeat." In: *Philosophy and Phenomenological Research* 88, pp. 314-45.

—— (forthcoming): "Enrasia or Evidentialism? Learning to Love Mismatch" Forthcoming in *Philosophical Studies*.

—— (this volume): "Defeat as Evincibility." In M. Skipper and A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, Oxford University Press.

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

—— (2015): "Stop Making Sense? On a Puzzle about Rationality." In: *Philosophy and Phenomenological Research*. Online First.

Pollock, J. (1974): *Knowledge and justification*. Princeton: Princeton University Press.

Rosenkranz, S. and M. Schulz (2015): "Peer Disagreement: A Call for the Revision of Prior Probabilities." In: *Dialectica* 69, pp. 551-86.

Schechter, J. (2013): "Rational Self-Doubt and the Failure of Closure." In: *Philosophical Studies* 163, pp. 428-52.

Schoenfield, M. (2015): "A Dilemma for Calibrationism." In: *Philosophy and Phenomenological Research* 91, pp. 425-55.

—— (2016): "An Accuracy Based Approach to Higher Order Evidence." In: *Philosophy and Phenomenological Research*.

Shah, N. (2006): "A New Argument for Evidentialism." In: *The Philosophical Quarterly* 56, pp. 481-98.

Silva, P. (2017): "How Doxastic Justification Helps Us Solve the Problem of Misleading Higher-Order Evidence." In: *Pacific Philosophical Quarterly* 98(1), pp. 308-28.

Skipper, M. (forthcoming): "Reconciling Enkrasia and Higher-Order Defeat." Forthcoming in *Erkenntnis*.

Steglich-Petersen, A. (2018): "Epistemic Instrumentalism, Permissibility, and Reasons for Belief." In: *Normativity: Epistemic and Practical*, Conor McHugh, J. Way and D. Whiting (eds.), Oxford: Oxford University Press, pp. 260-80.

—— (this volume): "Higher-Order Defeat and Doxastic Resilience." In M. Skipper and A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, Oxford University Press.

Titelbaum, M. (2015): "Rationality's Fixed Point (Or: In Defense of Right Reason)." In: *Oxford Studies in Epistemology* 5, T. Gendler and J. Hawthorne (eds.), pp. 253-94.

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

—— (2011): "Improbable Knowing." In: *Evidentialism and its Discontents*, T. Dougherty (ed.), Oxford: Oxford University Press.

—— (2014): "Very Improbable Knowing." In: *Erkenntnis* 79(5), pp. 971-99.

Worsnip, A. (2018): "The Conflict of Evidence and Coherence." In: *Philosophy and Phenomenological Research* 96(1), pp. 3-44.

—— (this volume): "How Your Total Evidence Can Mislead About Itself." In M. Skipper and A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, Oxford University Press.

Zhao, J., V. Crupi, K. Tentori, B. Fitelson, and D. Osherson (2012): "Updating: Learning versus Supposing." In: *Cognition* 124, pp. 373-78.