

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

Mattias Skipper

Draft of June 2018; final version to appear in M. Skipper and A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*. Oxford University Press.

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.

But a number of philosophers have recently 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 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 shall say more to clarify and motivate this idea in due course. 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 the next section, I begin by offering more precise formulations of evidentialism and the enkratic principle, before explaining why they jointly imply that a certain sort of self-misleading evidence is impossible. In §3, I then review what I take to be the most pressing challenge, due to Worsnip (2018), against the view that self-misleading evidence is impossible, and argue that the challenge is unconvincing as it stands. In §4, I proceed to outline and motivate a view of the normative significance 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 §5, I provide further motivation of the proposed view by showing how it may help explain a number of distinctive features of higher-order evidence that have been highlighted in the recent literature. Finally, in §6, I briefly sum up.

2. Evidentialism, Enkrasia, and Self-Misleading Evidence

Let me begin with a few remarks on terminology and notation. As usual, if p is a proposition, I will write ' Bp ' to say that p is believed by the relevant agent. 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: 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 for the operators r , R , e , and E , but formally inclined readers may think of them as pairs of possibility and necessity operators from standard epistemic and deontic logic, in the tradition of 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 convenience, 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. The term 'doxastic attitude' will be used to pick out a binary rather than a graded notion. Accordingly, I will be talking about (outright) belief, disbelief, and suspension of judgment, but I will not be talking about credences or levels of confidence. The reason for this choice is that the enkratic principle is traditionally understood in terms of coarse-grained "all-or-nothing" attitudes rather than in terms of graded ones. Of course, credal states can presumably also be akratic; but to keep matters relatively simple, I shall focus my attention on the traditional formulation of the enkratic principle in terms of binary attitudes.

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

Evidentialism: Necessarily,

(i) $ra \leftrightarrow ea$; and

(ii) $Ra \leftrightarrow Ea$.

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 this view 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, say, 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. In particular, someone might hold that even the strongest evidence only ever makes it rationally *permissible* (not *required*) to adopt this-or-that doxastic attitude towards a given proposition. That is, someone might accept the first clause in Evidentialism, but deny the second.¹ However, since both clauses will be needed to establish the conclusion that self-misleading evidence is impossible, I shall focus my attention on this relatively strong version of evidentialism.

How plausible is Evidentialism? It seems intuitively clear that a person’s evidence somehow plays an important role in determining which doxastic attitudes it is rational for the person to adopt. If we are asked to judge whether someone’s beliefs on some matter are rational, it seems perfectly appropriate to reflect on whether the person has responded in a reasonable manner to the evidence at hand. But despite its *prima facie* appeal, Evidentialism remains subject to ongoing debate (see, e.g., Conee and Feldman (2004), Littlejohn (2012), and Shah (2006) for critical discussions of evidentialism in different guises). Given the dialectics of the present paper, I do not want to enter a detailed discussion of the qualms some philosophers have had about posing a strong connection between evidential support and

¹ See, e.g., Conee and Feldman (1985) and Steglich-Petersen (2018).

epistemic rationality. Rather, I will simply take Evidentialism for granted in order to allow for the challenge, to which I aim to respond, to 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, one is 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, one is never permitted to believe that "I shouldn't believe that it's raining" while believing that "it's raining." Hence, in contrast to Evidentialism, the Enkratic Principle is a *structural* requirement of rationality in virtue of constraining which *combinations* of doxastic attitudes it may be rational to have. That is, the Enkratic Principle says nothing about which *particular* doxastic attitudes agent should adopt, but merely asks them to maintain a certain coherence between their doxastic attitudes and their beliefs about which doxastic attitudes they ought to have.

How plausible is the Enkratic Principle? At first blush, many seem to find the principle borderline self-evident. Just as it seems manifestly irrational to believe Moorean propositions like "it's raining, but I don't believe that it's raining" or contradictory propositions like "it's raining, but it isn't 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 not universally accepted among philosophers, but remains subject to ongoing debate.² However, as above, I shall not enter a general discussion of its merits or demerits. For present purposes, I will simply take the Enkratic Principle for granted alongside with Evidentialism.

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).

A number of authors have recently observed that a commitment to evidentialism leads to violations of 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 rule out certain types of self-misleading evidence as impossible. Given the above formulations of evidentialism and the enkratic principle, we can state this implication in a formally precise way (a proof can be found in the Appendix):

Impossibility of Self-Misleading Evidence: Necessarily,

- (i) $Ea \rightarrow \sim eB\sim ea$; and
- (ii) $\sim ea \rightarrow \sim eBEa$.

According to this thesis, 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, 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: evidence that sufficiently supports believing that it does not sufficiently support a doxastic attitude that it in fact decisively supports; and evidence that sufficiently supports believing that it decisively supports a doxastic attitude that it in fact does not sufficiently support.

At this point, let me pause to explain, in informal terms, why Evidentialism and the Enkratic Principle jointly imply the Impossibility of Self-Misleading Evidence. 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 a denial of the first clause in the Impossibility of Self-Misleading Evidence. If we

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.

furthermore assume that Evidentialism is true, it then follows that the agent in question is at once rationally required to adopt the attitude a and rationally permitted to believe that a is not sufficiently supported by the agent's evidence. Hence, the agent is at once rationally required to adopt a and rationally permitted to believe that she is not rationally permitted to adopt a .⁴ Yet, this conclusion 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 Impossibility in Self-Misleading Evidence can be established in a similar way using the first clause in the Enkratic Principle.

Although Evidentialism and the Enkratic Principle together rule out *certain* sorts of self-misleading evidence, it is worth noting that 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) $ea \ \& \ eB \sim ea$; or
- (ii) $\sim ea \ \& \ eBea$.

According to this thesis, it is indeed possible for certain false beliefs *about* what one's evidence supports to be supported *by* one's evidence. More precisely, according to the first clause, 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." Conversely, according to the second clause, one's evidence might at once not sufficiently support a , but sufficiently support believing that it does sufficiently support a . For example, my evidence might not sufficiently support believing that "it's raining", but sufficiently support believing that "my evidence does sufficiently support believing that it's raining." We can thus sum up the foregoing observations by saying that Evidentialism and the Enkratic Principle are compatible with the possibility of *weakly* self-misleading evidence, but incompatible with the possibility of *radically* self-misleading evidence. Of course, it might seem like an *ad hoc* or unstable position to maintain that weakly self-misleading evidence is possible, whereas radically self-misleading

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.

evidence is not.⁵ But in any case, it is worth being clear about what is a direct consequence of Evidentialism and the Enkratic Principle, and what is not.

The established connection between Evidentialism, the Enkratic Principle, and the Impossibility of Self-Misleading Evidence is going to constitute a ‘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 hope to show that the former option is a viable one: we *can* indeed reasonably deny the possibility of self-misleading evidence, and thereby put ourselves in a position to save both evidentialism and the enkratic principle. But first, I want to 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 his evidence: he has many years of experience, is rarely on the phone whilst driving, 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 formula T is tautological or not. As it happens, Sophie makes a few errors and draws the wrong conclusion that T is non-tautological. All of Sophie’s classmates happen to reach the same wrong conclusion. Sophie is aware of their agreement. What’s more, the 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 begin by taking a closer look at the evidence in each case.

⁵ Although see Dorst (ms.) for an interesting proposal along these lines.

In Driver's Bias, John starts out with a body of evidence consisting of information about his own driving history, which is assumed to strongly indicate that his driving skills are above average. This kind of evidence is sometimes said to be of 'first order', because it seems *directly* relevant to the question of whether John is better than average at driving. Thus, by assumption, John's first-order evidence strongly suggests that he is better than average at driving. However, John then receives some additional evidence, namely the information about the driver's bias, which indicates that he has overestimated his own driving skills. This kind of evidence is sometimes said to be of 'higher order', because it only seems *indirectly* relevant to the question of whether John's driving skills are above average (in virtue of being directly relevant to the question of whether John has overestimated his own driving skills). I shall later (in §5.4) offer a more precise characterization of this intuitive distinction between 'direct' and 'indirect' evidential bearing as it applies in the context of distinguishing between first-order and higher-order evidence. But for now, it suffices to have a merely case-based understanding of the distinction.

Similarly, 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 relevant 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; and, 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*.⁶

How strong are these seemingly opposing evidential relations in Driver's Bias and Poor Logic? Given the dialectical setting, 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 his belief that he is better than average at driving, and let us suppose that his higher-order evidence is at least

⁶ 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.

strong enough to sufficiently support believing that his first-order evidence does not sufficiently support believing that he is better than average at driving. Likewise, let us suppose that Sophie's first-order evidence does not sufficiently support her belief that T is non-tautological, and let us suppose that her higher-order evidence at least sufficiently supports believing that her first-order evidence decisively supports believing that T is non-tautological.

We can state these four stipulations about the evidential relations in Driver's Bias and Poor Logic more conveniently by introducing the following shorthands for the propositions featuring 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 .

Formulated in terms of these abbreviations, our four stipulations become: John's first-order evidence decisively supports believing P_F ; John's higher-order evidence sufficiently supports believing $\sim P_H$; Sophie's first-order evidence does not sufficiently support believing Q_F ; and Sophie's higher-order evidence sufficiently supports believing Q_H .

We can 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_E ; 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, self-misleading evidence is possible, which would force us 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'): a body of evidence need not support a doxastic attitude just because a subset of the evidence does. Countless examples can be found in the literature on epistemic defeat, and I shall not extend the list here.⁷ The point I want to make is simply that it would be too hasty to conclude that Driver's Bias and Poor Logic are cases of self-misleading evidence on the basis of considerations about what the respective bodies of first-order and higher-order evidence support when considered separately. What matters is what the first-order and higher-order evidence supports when considered in conjunction. This much should be uncontroversial.

But in a recent paper, Worsnip (2018) puts forth what may be seen as a refined version of the naïve argument, which purports to show that even if we take into account the possibility that Driver's Bias and Poor Logic are cases of epistemic defeat, they still constitute genuine counterexamples to the Impossibility of Self-Misleading Evidence (at least on *some* ways of filling in the details of the cases). Here is a minimal and somewhat liberal reconstruction of Worsnip's argument, as it applies to Driver's Bias:⁸

Asymmetry argument

- W₁** John's first-order evidence supports P_F more strongly than his higher-order evidence supports $\sim P_F$.
- W₂** John's higher-order evidence supports $\sim P_H$ more strongly than his first-order evidence supports P_H .
- W₃** So, John's total evidence supports both P_F and $\sim P_H$.

⁷ For overviews of different issues related to epistemic defeat, see Kelly (2016) and Koons (2017).

⁸ 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 particular case of misleading higher-order evidence is used as a backdrop for the present discussion.

Two initial remarks about this reconstruction are in order. First, the exposition in Worsnip (2018) is in large part 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 P_H . However, since these claims have no straightforward bearing on the relevant conclusion, which is W_3 , I take it that Worsnip ultimately wants (and needs) to establish two asymmetry claims akin to W_1 and W_2 that *do* bear on W_3 . Second, the conclusion W_3 is obviously 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. Rather, I want to grant that W_3 , if true, indeed constitutes a genuine counterexample to the Impossibility of Self-Misleading Evidence. For similar dialectical reasons, I will also straightaway concede the asymmetry claims W_1 and W_2 .

Still, 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 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: just as the naïve argument draws a conclusion about what the total evidence supports from premises about what the first-order 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 and higher-order evidence supports when taken separately.

Might there be a way of modifying the Asymmetry argument to avoid this problem? One could introduce some additional premises that would ensure, roughly speaking, that the presence of the first-order evidence does not significantly alter the bearing of the higher-order evidence on P_F and P_H , and that the presence of the higher-order evidence does not significantly alter 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 .

The term ‘undercutting’ is here used in the customary way, as introduced by Pollock (1974). That is, 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. So, if we conjoin 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 good reason to reject No Top-Down Undercutting; or so I shall argue. Even so, one might obviously 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 shall offer against No Top-Down Undercutting are going to speak equally against such weakened versions of that principle. Hence, if the considerations below are ultimately right, 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

Thus far, I have deliberately avoided to 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 a dialectical one, namely that I think Worsnip’s challenge to the Impossibility of Self-Misleading Evidence can be resisted without any appeal to intuition. 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, and that it sufficiently (if not decisively) supports believing that his first-order evidence does not sufficiently support believing that his driving skills are above average. Likewise, I take it to be intuitive that Sophie’s total evidence sufficiently supports believing that T is non-tautological, and that it sufficiently (if not decisively) supports believing that her first-order evidence sufficiently supports believing that T is non-tautological. For what they are worth, these intuitive verdicts suggest that Driver’s Bias and Poor Logic are *not* cases of self-misleading evidence: if John’s total evidence indeed supports $\sim P_H$ but not P_F , while Sophie’s total evidence supports both Q_F and Q_H , neither body of

evidence is self-misleading. Needless to say, it is a matter of controversy how much weight we should place on such intuitive judgments in the context of justifying our philosophical theories. I take it that our judgments in the cases at hand are robust enough to carry at least *some* argumentative weight. But more importantly, I also think we can make good *theoretical* sense of them. Indeed, as I shall argue below, there is an independently plausible view of higher-order evidence available that at once vindicates our intuitive verdicts in cases like Driver's Bias and Poor Logic, and 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 David Christensen (2010; 2011), of the underlying theoretical commitments that give 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 are to maintain in light of the higher-order evidence that their original beliefs are supported by their first-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 . Something similar holds of Sophie: just as John cannot reasonably *continue* to maintain that his first-order evidence supports P_F , Sophie can reasonably *begin* to maintain that her first-order evidence supports Q_F in light of the higher-order evidence.

We can spell out these considerations a bit 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, what the above considerations suggest is that it would be irrational of him to consider P_F likely on the first-order evidence *after* having received the higher-order evidence. If he *did*, he would thereby disregard the higher-order evidence in what looks like a dogmatic manner. Something similar holds of Sophie: she should not consider Q_F likely on the

first-order evidence *before* receiving the higher-order evidence, but may reasonably consider Q_F likely on the first-order evidence *after* having received the higher-order evidence.

In light of these considerations, it looks like 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 should not be significantly altered by the presence of John's higher-order evidence, and the evidential relation between Sophie's first-order evidence and Q_F should likewise not be significantly altered by the presence of her higher-order evidence. Yet, this is precisely what we have denied on the grounds that John and Sophie would otherwise end up assimilating their respective bodies of higher-order evidence in what appears to be an unreasonable manner. Thus, to the extent that the kind of dogmatism identified by Christensen is indeed irrational, the prospects look dim for No Top-Down Undercutting. Consequently, we have a diagnosis of where, exactly, the Asymmetry argument against the Impossibility of Self-Misleading Evidence goes wrong: as we recall from the previous section, the inference from W_1 and W_2 to W_3 presupposes two additional assumptions along the lines of No Top-Down Undercutting and No Bottom-Up Undercutting. So if No Top-Down Undercutting must be rejected, the Asymmetry argument cannot ultimately be made to work.

But the present considerations against No Top-Down Undercutting do not merely fend off the Asymmetry argument; they also lend more direct support to the view that self-misleading evidence is impossible. To see why, it is useful to reformulate the foregoing points about the normative impact of John and Sophie's respective bodies of higher-order evidence in terms of the changes 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 wet streets to be strong evidence for recent rain. By contrast, my doxastic attitude towards "it has recently been raining" conditional on "the streets are dry" is *disbelief*, since I take dry streets 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.⁹ Hence, I take the notion of a conditional *all-or-nothing* doxastic attitude to play much the same role in the present context as the notion of a conditional *graded* doxastic attitude plays in a Bayesian context. Just as conditional graded doxastic attitudes reflect the way in which Bayesian agents take different bodies of evidence to bear on different propositions, so conditional all-or-nothing doxastic attitudes reflect the way in which ‘non-Bayesian’ agents take different bodies of evidence to bear on different propositions.¹⁰

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? Recall that before John receives the relevant higher-order evidence, his doxastic attitude towards P_F conditional on the first-order evidence is belief. By assumption, this conditional doxastic attitude 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 have his original conditional doxastic attitude towards P_F , since the higher-order evidence indicates that his first-order evidence does not support P_F . That is, 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 may be told about Sophie: before she receives the relevant higher-order evidence, her doxastic attitude towards Q_F conditional on the first-order evidence is belief. By assumption, this conditional doxastic attitude is irrational, since her first-order evidence is assumed not to support Q_F . However, after having received the higher-order evidence, it is rational for Sophie to have her original conditional doxastic attitude 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 doxastic attitude, Sophie’s higher-order evidence alleviates an existing rational pressure to revise her original conditional doxastic attitude.

⁹ 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.

¹⁰ This obviously raises some further questions about how an agent’s *conditional* all-or-nothing doxastic attitudes should relate to the agent’s *unconditional* all-or-nothing doxastic attitudes. Perhaps we might try to constrain conditional all-or-nothing doxastic attitudes in terms of a binary analogue to the Ratio Formula that is standardly used by Bayesians to relate an agent’s conditional credences to the agent’s unconditional credences. However, we need not settle these details here. For present purposes, it suffices to have a basic informal grasp of the notion of a conditional all-or-nothing doxastic attitude.

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 attitudes it is rational to have*. In Driver’s Bias, the higher-order evidence has a normative impact on John’s doxastic attitude towards P_F , because it requires him to revise his conditional doxastic attitude towards P_F from belief to suspension of judgment. In Poor Logic, the higher-order evidence has a normative impact on Sophie’s doxastic attitude towards Q_F , because it makes it rational for her to retain her conditional doxastic attitude towards Q_F . We thus 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, nor do other structurally similar cases of misleading higher-order evidence.

Obviously, none of this conclusively shows that self-misleading evidence can *never* occur. Even if cases like Driver’s Bias and Poor Logic are not cases of self-misleading evidence, there might be *other* types of cases that *are*. 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 one’s evidence (even if one cannot be misled about what one’s evidence *supports*).¹¹ This issue is firmly rooted in the ongoing dispute between internalists and externalists about epistemic rationality, and lies well beyond the scope of this paper. But at the very least, 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 I proceed, let me note that the idea that we sometimes ought to revise our conditional doxastic attitudes is not a new one. For instance, Lange (1999) has argued that a number of central problems in Bayesian confirmation theory can be resolved if we allow for certain credal changes to come about as the result of revising our conditional credences (or ‘confirmation commitments’ as he follows Levi (1980) in calling them). Along similar lines,

¹¹ 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.

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. So the idea 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 want to suggest here is that the same idea may serve as the basis of a fairly general view of the normative significance of higher-order evidence.

5. Further Explanatory Attractions

With the core proposal on the table, I now want to provide some further motivation of the idea that higher-order evidence works by influencing which conditional doxastic attitudes it is rational to have. More specifically, the aim in the following is to illustrate how this idea may help to explain various distinctive features of higher-order evidence that have been pointed out in the recent literature. Doing so will also give me an opportunity to clarify and elaborate a bit on different aspects of the proposed view.

5.1 *The Retrospectivity of Higher-Order Evidence*

First off, the proposed view can help to explain the *retrospective* aspect of higher-order evidence. A number of authors have pointed out that 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.¹² Take again John as an example: when he receives the information about the driver's bias, he thereby acquires a reason to think that his belief in P_F was never supported by his first-order evidence *to begin with*.

By contrast, first-order evidence does not display the same sort of retrospectivity. If, for example, I believe that it is raining outside on basis of the testimony of a reliable friend, but then look out the window to 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. Or, if I believe that the wall in front of me is red on the basis of its looking red, but then learn that the wall is merely lit up by a red spotlight, I do not thereby acquire 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 a kind of retrospectivity that first-order defeaters do not? The proposed view of higher-order evidence offers a simple explanation: the reason

¹² See, e.g., Christensen (2010), Lasonen-Aarnio (2014), and DiPaolo (2016).

why higher-order defeaters have a retrospective aspect is that they work by indicating that one's original conditional doxastic attitude (that is, the conditional doxastic attitude one had before receiving the higher-order defeater) was initially irrational. By contrast, first-order defeaters do not have a retrospective aspect, because they do *not* work by indicating that one's original conditional doxastic attitude was initially irrational. On the present picture, this is why higher-order evidence and first-order evidence differ in their retrospective character.

5.2 *The Agent-Relativity of Higher-Order Evidence*

Secondly, the proposed view can help to explain why the normative force of a body of higher-order evidence in many cases depends on *who possesses it*. To illustrate this agent-relativity, consider a case adapted from Christensen (2010, p. 187):

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

In this case, it looks like the higher-order evidence (that is, the information that I have been given a reason-distorting drug) has very different implications for how confident You and I should end up being that the square root of 121 is indeed 11: while I should lose confidence in my belief, at least to some extent, there is no apparent reason for You to do the same. After all, the fact that someone else than yourself is drugged is hardly relevant to the question of whether the square root of 121 equals 11.

As pointed out by, e.g., Christensen (2010) and Kelly (2005), this sort of agent-relativity is a rather peculiar feature of higher-order evidence. We are used to think that the question of *how* an agent should respond to a given body of evidence should not depend for its answer on *who* the agent is. Of course, the answer might depend on the agent's background information, since evidential support is not in general monotonic. But the kind of agent-relativity that is at stake in Arithmetic on Drugs remains in place even if we assume that You and I have exactly 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.

Why is the normative significance of higher-order evidence agent-relative in this way? Again, the proposed view offers a simple explanation: in Arithmetic on Drugs, the central difference between You and I 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 “ $\sqrt{121} = 11$.” Consequently, I should revise my conditional doxastic attitude towards “ $\sqrt{121} = 11$,” whereas You should not revise your conditional doxastic attitude towards “ $\sqrt{121} = 11$ ”. On the present picture, this is why the same body of higher-order evidence may have very different normative implications for different agents, even if those agents share the same background information.

5.3 *The Insignificance of Hypothetical Higher-Order Evidence*

Third, the proposed view can help to explain why 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 will carry out an experimental test of a hypothesis *h* during next week. The experiment has two possible outcomes, o_1 and o_2 , where o_1 would strongly confirm *h* and 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 a sound assessment of the experimental results.

Let us follow Christensen in asking two questions about Brenda. First, how confident should she *now* be in *h* conditional on o_1 in conjunction with the fact that she will be cognitively impaired next week? Intuitively, it seems, very confident! After all, the fact that she will be cognitively impaired next week is hardly relevant to the question of whether o_1 supports *h* or not. Second, if Brenda actually learns o_1 next week, how confident should she *then* be that *h* is true? Intuitively, it seems, *not* very confident! After all, she will in that case be aware of being cognitively impaired in a way that makes her unable to give a sound assessment of the experimental results. Yet, if we put these two verdicts together, we end with the result that Brenda’s doxastic attitude towards *h* next week should not match her current doxastic attitude towards *h* conditional on o_1 in conjunction with the fact that she will be cognitively impaired during next week.

This is a striking result. We are used to think that an agent's doxastic attitudes after having acquired a body of evidence should match the agent's (relevant) conditional doxastic attitudes before having acquired that evidence. In other words, we are used to think 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. Indeed, 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 idea with a standard case of first-order defeat:

Defective Experiment: Joe is a formidable scientist who will carry out an experimental test of a hypothesis h during next week. The experiment has two possible outcomes, o_1 and o_2 , where o_1 would strongly confirm h and o_2 would strongly disconfirm h . Being a formidable scientist, Joe is well aware of these evidential relations. But when 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 in conjunction with the fact that one of the key measurement devices will be defective next week? Intuitively, it seems, not very confident! After all, the defective instrument would render the entire experiment defective. 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 ? As before, it seems, not very unconfident! That is, it seems that Joe's doxastic attitude towards h after having conducted the experiment should match his current doxastic attitude towards h conditional on o_1 in conjunction with 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 depend on whether it is learnt or supposed.

Why does higher-order evidence differ from first-order evidence in this respect? The proposed view once again offers a possible explanation: when Brenda learns that she is drugged, she should not simply respond to this higher-order evidence by conditionalizing on it, but should also revise her doxastic attitude towards h conditional on o_1 (from belief to

suspension of judgment in the case at hand). Consequently, her resulting doxastic attitude towards h will not match her prior doxastic attitude towards h conditional on o_i in conjunction with 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. Consequently, his resulting doxastic attitude towards h will match his prior doxastic attitude towards h conditional on o_i in conjunction with the fact that one of the key measurement devices is defective next week. On the present picture, this is why higher-order evidence (at least sometimes) depends 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*

Finally, the proposed view can help to make sense of the previously mentioned intuition that first-order evidence somehow bears *directly* on the proposition at hand in a way that higher-order evidence does *not*. This distinction between ‘direct’ and ‘indirect’ evidence is somehow supposed to capture two distinct ways in which a body of evidence might bear on a proposition. But what, exactly, might this supposed difference amount to?

The proposed view suggests the following answer: higher-order evidence is ‘indirect’ evidence in the sense that it influences which doxastic attitudes it is rational to have *by way of* influencing which conditional doxastic attitudes it is rational to have (conditional, that is, on the relevant first-order evidence). For example, John’s higher-order evidence about the driver’s bias bears indirectly on the question of whether he is a superior driver in virtue of requiring him to revise his doxastic attitude towards P_F conditional on the first-order evidence about his driving habits and track-record. By contrast, first-order evidence is ‘direct’ evidence in the sense that it influences which doxastic attitudes it is rational to have *by way of* requiring one to conditionalize one’s doxastic attitudes on it. For example, John’s first-order evidence bears directly on the question of whether he is a superior driver in virtue of 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 suggested that higher-order evidence acquires its normative force by influencing which conditional doxastic attitudes it is rational to have. If so, we have a way of maintaining that

cases of misleading higher-order evidence about what one's first-order evidence supports are not cases of self-misleading evidence. This is an important result, since Evidentialism and the Enkratic Principle jointly imply that such self-misleading evidence is impossible. Thus, the considerations put forth give us a way of accommodating the Enkratic Principle within an evidentialist framework.

This obviously does not show that either thesis is true. Even if what I have said about the normative significance of higher-order evidence is basically correct, there might be independent reasons to reject Evidentialism or the Enkratic Principle (or both). But I hope to have provided at least some tentative reasons to think that Evidentialism, the Enkratic Principle, and the proposed view of higher-order evidence together form an attractive theoretical package.

The present chapter is clearly just a first step towards a complete theory of the normative significance of higher-order evidence. This is partly due to the lack of a detailed comparison with other candidate approaches to higher-order evidence, which I have not aimed to undertake in this chapter.¹³ But perhaps more to the point, I have done little to spell out *how*, exactly, a given body of higher-order evidence should influence a given conditional doxastic attitude that an agent might have. I suspect that the various peculiar features of higher-order evidence discussed in the previous section will feature as central pieces in the final puzzle. But the finer details must wait for future work.

Acknowledgements. An earlier version of this paper was present in the Department of Philosophy at the University of Cologne. I would like to thank the audience on that occasion for very helpful feedback. Thanks also to Alex Worsnip, Asbjørn Steglich-Petersen, and Jens Christian Bjerring for extremely helpful comments and criticism.

Appendix

In this appendix, I show that Evidentialism and the Enkratic Principle jointly entail the Impossibility of Self-Misleading Evidence. Let ' \Box ' represent conceptual necessity, and let ' \Diamond ' be its dual. The following auxiliary assumption is needed for the argument to go through:

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

¹³ For example, Silva (2017) proposes to understand higher-order defeat in terms of loss of 'doxastic justification', whereas Steglich-Petersen (this volume) proposes to understand higher-order defeat in terms of loss of 'doxastic resilience'.

According to rB -closure, one is rationally permitted to believe the necessary consequences of what one is rationally permitted to believe. For example, if one is rationally permitted to believe that “it’s raining, and it’s only raining if the streets are wet,” one is rationally permitted to believe that “the streets are wet.” Of course, this is not an uncontroversial assumption, but given the dialectics of the paper, we can grant it to allow for the inference of the Impossibility of Self-Misleading Evidence to go through.¹⁴

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)
5. $\Diamond r(a \& B\sim ra)$ (4)
6. $\Diamond \sim R(B\sim ra \rightarrow \sim a)$ (5)
7. $\Box(Ea \rightarrow \sim eB\sim ea)$ (reductio 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. 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.

¹⁴ For further discussion of epistemic closure principles, see Klein (1981; 1995) and Schechter (2013).

- (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.”
- (ms.): “Evidence: A Guide for the Uncertain.” Unpublished manuscript.
- 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: *Phil. Studies* 167, pp. 201-19.
- 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: 213–236.

- Koons, R. (2017): "Defeasible Reasoning". In: *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): "Enkrasia or Evidentialism? Learning to Love Mismatch" In: *Philosophical Studies*.
- (this volume): "Defeat as Evincibility"
- 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". In: *Erkenntnis*.
- Steglich-Petersen, A. (2018): "Epistemic Instrumentalism, Permissibility, and Reasons for Belief". In: *Normativity: Epistemic and Practical*, Conor McHugh, J. Way & D. Whiting (eds.), Oxford: Oxford University Press, pp. 260-80.
- (this volume): "Higher-Order Defeat and Doxastic Resilience"
- 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.”
- Zhao J., Crupi V., Tentori K., Fitelson B., and Osherson D. (2012): “Updating: Learning versus Supposing.” In: *Cognition* 124, pp. 373-78.