DETACHING IN PEER DISAGREEMENT AND HIGHER-ORDER EVIDENCE

by

Eyal Tal

_____

Copyright © Eyal Tal 2018

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements For the Degree of

DOCTOR OF PHILOSOPHY

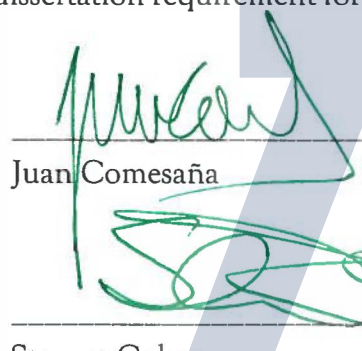
In the Graduate College

THE UNIVERSITY OF ARIZONA

2018

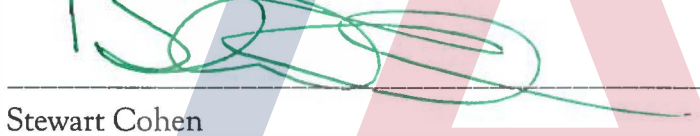
THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Eyal Tal, titled 'Detaching in Peer Disagreement and Higher-Order Evidence' and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.



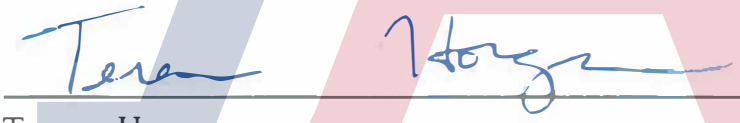
Date: (04/06/2018)

Juan Comesaña



Date: (04/06/2018)

Stewart Cohen

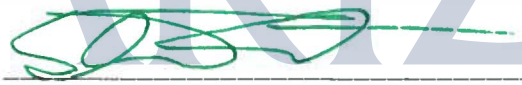


Date: (04/06/2018)

Terence Horgan

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.



Date: (04/06/2018)

Dissertation Director: Stewart Cohen

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Eyal Tal

ACKNOWLEDGEMENTS

Much of this work defends the position that it is rational, although impractical, to dismiss our epistemic peers and superiors. Fortunately, I am a practical person. I wish to thank Stewart Cohen, Juan Comesaña, and Terence Horgan for offering invaluable advice and guidance throughout my graduate career. Their generosity, patience, honesty, and encouragement, are responsible for my producing this work.

CONTENTS

ABSTRACTS

In this three-paper project I discuss the detaching problem for principles about the rational response to higher-order evidence. In the first paper, I argue that the problem presents a more serious problem than previously thought to conciliatory principles. In the second paper I argue that the problem applies to a wide variety of potential principles regarding the rational response to higher-order evidence. In the third paper, I argue that the problem suggests that the rational response to higher-order evidence is no response at all, i.e., that it is rational to dismiss higher-order evidence.

Paper 1. Should conciliating with disagreeing peers be considered a sufficient condition for reaching a rational belief? Thomas Kelly argues that when taken this way, Conciliationism allows for all too easy acquisition of rational beliefs. Two responses defending Conciliationism have been offered. One response views conciliation as sufficient for holding a rational belief, but only requires it of agents who enter into a disagreement with a rational belief. This response makes for a requirement that no one should follow. If the need to conciliate only applies to already rational agents, then an agent must conciliate only when her peer is the one irrational. The other response views conciliation as merely necessary for holding a rational belief. This response does not answer the central question of what is rational to believe when facing a disagreeing peer. Attempts to develop this response either collapse into the first response or give rise to frequent rational dilemmas.

Paper 2. Certain attitudes come with strings attached. For instance, believing that chocolate is harmful to dogs commits us to not feeding our dogs chocolate. Focusing on doxastic attitudes, epistemologists habitually offer principles that aim to capture what beliefs we are committed to adopting or avoiding given other beliefs that we have. A

known worry with some of these principles is that they seem to open the door to troublesome rational dilemmas. If believing ~P commits us to disbelieving P, and we believe ~P despite evidence to the contrary, then our commitment would have us disbelieve something that our evidence requires we believe. Attempts to avoid the problem share the result that only a rationally held belief commits its holder to adopting or avoiding other beliefs. This result has a surprising implication for possible principles regarding the rational response to higher-order evidence. If higher-order evidence requires belief revision only of agents who responded rationally to their original evidence, then higher-order evidence requires belief revision only when it is misleading. It would be perfectly rational of agents who notice this never to adhere to those principles, knowing that they do nothing but mislead. As a result, a variety of plausible-sounding principles regarding the rational response to higher-order evidence are incorrect.

Paper 3. Suppose we learn that we have a poor track record in forming beliefs rationally, or that a brilliant colleague thinks that we believe P irrationally. Does gaining such information require us to revise those beliefs whose rationality is questioned? When we gain information suggesting that our beliefs are irrational, one of two general cases obtains. In the first case we made no mistake and our beliefs are rational. In that case the information to the contrary is misleading. In the second case we indeed believe irrationally, and our original evidence *already* requires us to revise our belief. In that case, information to that effect is superfluous. Thus, information to the effect that our beliefs are irrational is either misleading or superfluous, and cannot justify belief revision.

Disagreement and Easy Bootstrapping: Does Conciliationism Apply to No One?

Eyal Tal

University of Arizona

Abstract: Should conciliating with disagreeing peers be considered a sufficient condition for reaching a rational belief? Thomas Kelly argues that when taken this way, Conciliationism allows for all too easy acquisition of rational beliefs. Two responses defending Conciliationism have been offered. One response views conciliation as sufficient for holding a rational belief, but only requires it of agents who enter into a disagreement with a rational belief. This response makes for a requirement that no one should follow. If the need to conciliate only applies to already rational agents, then an agent must conciliate only when her peer is the one irrational. The other response views conciliation as merely necessary for holding a rational belief. This response does not answer the central question of what is rational to believe when facing a disagreeing peer. Attempts to develop this response either collapse into the first response or give rise to frequent rational dilemmas.

1 Introduction: Conciliationism and Easy Bootstrapping

According to Conciliationism, one must lose confidence in one's view upon meeting a dissenting epistemic peer.[1] The further a peer's view is from one's own, the more confidence one must lose. On a graded framework for doxastic attitudes, many understand

[1] Many authors have, at some point or another, defended versions of this view. See Feldman (2007), Christensen (2007), Elga (2007) and Cohen (2013).

Conciliationism as requiring each peer to revise her credence in the direction of the other.

Thomas Kelly (2010, 2014) worries that Conciliationism makes it too easy to form fully rational credences.[2] The argument follows a simple recipe. First, take an agent S who assigns an irrational credence to some proposition P. Then, introduce S to a peer who disagrees. Finally, per Conciliationism, have S conciliate to the required degree. And there you have it: S now holds a rational credence in P. The problem is most vivid when both agents start out irrational, and each ends up with a rational credence by conciliating with the other. Following Kelly, let us call this the *Easy Bootstrapping* argument (EB).[3] EB shows that an unrestricted version of Conciliationism must be wrong.

In response to EB, philosophers have offered versions of Conciliationism that aim to render it immune to the argument. These versions either restrict Conciliationism so that irrational agents are not required to conciliate, or interpret the view as providing merely a necessary condition for rational belief, rather than a sufficient one. I argue that both kinds of response are unpromising. Restricting the need to conciliate to agents who hold rational credences would ensure that following Conciliationism would lead those rational agents away from the rational credence they hold on the shared evidence. Importantly, added evidence from the disagreement would not give those agents reason to revise their rational credences, because their disagreeing peers would always be holding irrational credences. Thus, the restricting move results in a conciliatory requirement that is rational not to

---

2 For his original argument see Kelly (2010), pp. 126–7.

3 Weatherson (ms.) names this phenomenon *epistemic laundering*.

follow. The alternative move of interpreting Conciliationism merely as a necessary condition for arriving at a rational credence leads to one of three unwanted results: either it does not tell us what to believe, or it collapses into the restricting response, or it gives rise to conflicting rational requirements.

I start the discussion by showing that EB is an instance of the infamous detaching problem, which many otherwise attractive normative principles face. Noticing this helps to see both why the voiced responses to EB seem viable initially, and also why some of them ultimately fail.

2 Easy Bootstrapping and the Detaching Problem

It has gone unnoticed that EB is the result of applying the detaching problem to Conciliationism. Stephen Finlay (2010) illustrates the detaching problem using the following principle:

Self-Reliance: If an agent S believes that she ought to φ, then she ought to φ.

Finlay points out that, on the one hand, a lot seems right about Self-Reliance. There is something amiss with agents who believe they ought to φ but do not φ. On the other hand, the principle yields an unintuitive result, namely, that we can never falsely believe we ought to φ. But surely this is wrong. Merely believing we should φ does not make it the case that we should φ. For example, irrationally believing we should assassinate the mayor for fun

does not make it the case that we should. More generally, the problem stems from the fact that the consequent of a hypothetical imperative detaches *whenever* its antecedent is satisfied.[4] In those cases where the antecedent is true due to the agent's irrationality, some hypothetical imperatives issue dubious requirements.

The requirements issued by principles susceptible to a detaching problem appear, in certain cases, obviously incorrect. But even if we do not find the requirements to be obviously incorrect, there is other reason for concern. If a principle's detached consequent requires something that conflicts with other, intuitively correct requirements, the principle would be responsible for normative dilemmas. For example, a principle saying that *if we drive drunk we should drive with our hazards on* does not seem to issue an obviously incorrect recommendation for drunk drivers, even though that recommendation conflicts with the intuitively correct recommendation that they not drive drunk altogether. Since it is doubtful that genuine normative conflicts arise so frequently, alleged principles that give rise to them are suspect.

For certain normative principles the detaching problem spells further trouble in the form of a bootstrapping concern. If an agent's mere belief that she ought to φ makes it the case that she indeed ought to φ, then that seems like an overly easy way for her to ensure that she is doing the right thing. This point underlies the easy bootstrapping accusation that Kelly levels against Conciliationism.

---

[4] What I aim to capture here are conditional ought-statements in which the normative operator takes narrow scope. See Broome (1999).

Conciliationism can plausibly be construed as a principle that is structurally similar to those, like Self-Reliance, that face a detaching problem. A natural understanding of the conciliatory view gives us just such a principle:

> *Natural Conciliationism:* If one believes P to degree d_1, and one encounters a peer believing P to a different degree d_2, then rationality requires one to conciliate to an intermediate degree d_3.

The Easy Bootstrapping argument works against Natural Conciliationism in the same way the detaching problem works against Self-Reliance. When the antecedent of Natural Conciliationism is true due to the agent's irrationality, it seems wrong to say that the agent is thereby rationally required to conciliate. If the agent were indeed required to do so, she would seem to acquire a fully rational belief upon conciliating, despite the fact that her original belief was irrational. As Kelly notes, rational beliefs are not so easily had.

The offered responses to EB correspond to two kinds of responses to the detaching problem. The first kind restricts the relevant principle to agents who rationally satisfy the antecedent. Applied to Self-Reliance, such a restricted principle would say that *if you rationally believe that you ought to φ then you ought to φ.* The second kind of response construes the relevant principle as stating only a necessary condition for believing rationally. On one version of this line, Self-Reliance says that *it is necessary for doing the right thing that one's actions do not conflict with what one believes to be required.* I now go on to show that, for

reasons unique to peer disagreement, the first response, viz., restricting the principle's antecedent, does not succeed as a defense of Conciliationism. The second response, I then argue, either collapses into the restricting response, or fails to tell us what to believe in cases of peer disagreement, or admits of frequent normative dilemmas.

3 The Restricting Response

While discussing a particular version of Conciliationism—the Equal Weight View (EW)—Brian Weatherson (ms.) proposes we restrict the view in order to sidestep EB:

> Consider an agent who makes an irrational judgment. And assume her friend, who she knows to be a peer, makes the same irrational judgment. What does the Equal Weight View say she should do? It should be bad for it to say that she should regard her and her friend as equally likely to be right, so she should keep this judgment. After all, it was irrational! There are a couple of moves the friend of the Equal Weight View can make at this point. But I think the simplest one will be to put some kind of restriction on Equal Weight. [...] that restriction is to agents who have initially made rational judgments... (Weatherson ms., p. 5)

Weatherson suggests that we understand Conciliationism to require conciliating with a peer only when our initial credence is rational. With this restriction, when we start out irrational we are not required to conciliate. So the restriction successfully blocks EB. If we are at an irrational credence we are not guaranteed to reach a rational credence simply by

conciliating.

A conciliatory rule based on Weatherson's suggestion would look something like the following:

> *Restricted Conciliationism:* If one rationally believes P to degree d_1, and one encounters a peer believing P to a different degree d_2, then rationality requires one to conciliate to an intermediate degree d_3.

On this rule, conciliating in response to disagreement is both sufficient and necessary for reaching a fully rational belief state, but only for those who are already holding a rational credence. Since Restricted Conciliationism does not bind agents who enter into a disagreement with an irrational credence, such agents would not necessarily reach a rational credence by conciliating.

The move to restrict a normative rule only to cases where the antecedent is rationally satisfied seems promising. It works well as a way out of the detaching problem for similar rules, including Self-Reliance. Yet after originally raising EB, Kelly (2010) claims that such a move will not work, dismissing a response in the spirit of Weatherson's as "poorly motivated in the extreme."[5] Here is Kelly:

> If the phenomenon of peer disagreement requires you [the rational] to split the

[5] This appears in a footnote. See Kelly (2010), p. 128.

> difference with my unreasonable opinion, why should I be spared having to split the difference with your reasonable opinion simply in virtue of having botched the evidence in the first place? Whatever normative pressure is created by the phenomenon of peer disagreement, surely one does not immunize oneself against that pressure simply in virtue of having beliefs that are not adequately supported by one's evidence. (Kelly 2010, p. 128, my brackets)

But the restricting response to EB is not so easily dismissed. The response is not committed to disagreement putting no pressure on the irrational agent. Rather, it is silent about how irrational agents should revise, focusing only on what rational agents need to do to maintain their rationality. One could accept the restricted version of Conciliationism and also maintain that disagreement puts additional pressure on an irrational agent to revise, beyond the pressure that such an agent is already under from her pre-disagreement evidence.

Stewart Cohen (2013) is sympathetic to Weatherson's version of Conciliationism, but hopes for a more inclusive conciliatory view. Cohen entertains a way for initially irrational agents to fall under the conciliatory requirement:

> It may be that EW can be defended only as a theory of how one should revise when one is at a rational credence… Having said that, there may be a way for EW to explain how I [the irrational] end up with a rational credence at .5 [by conciliating]. We are assuming that when I learn of Peer's credence, I should treat it as evidence

> against the rationality of my credence. And that is because by stipulation, I have reason to think Peer is generally rational. Then by definition (of 'peer'), I have equally good reason to think I'm generally rational. That counts as (defeasible) evidence in favor of the rationality of my credence. (Cohen 2013, pp. 111-2, my brackets)

Cohen is after an ambitious conciliatory requirement, which would apply to all agents in peer disagreement situations, directing them to a fully rational credence. According to the argument, this can be done if the initially irrational agent reflects on her general rationality in the disputed domain. By stipulation, the agent is justified in giving her disagreeing interlocutor significant evidential weight, and this implies that the agent is also justified in taking herself to be generally rational.[6] If the agent reflects on her being generally rational, she could infer that her judgments are likely to be rational on that particular occasion. This, Cohen says, can make the agent justified in holding her original view (as long as her judgment is not *obviously* irrational) at which point conciliating would suffice for her to reach a fully rational credence. Call this the *General Rationality* argument.

There are places to resist the General Rationality argument. For example, Kelly (2014) argues that it allows for "even easier, more implausible bootstrapping." If knowing we are generally rational can allow us to upgrade our irrational beliefs to rational ones by reflecting on our general rationality, then rational belief would seem to come rather cheap.

[6] This follows from defining 'peers' as not just of equal rational standing, but also of high standing.

But resisting the General Rationality argument is not necessary just yet. This is because the argument fails to show how an agent holding an irrational credence may *conciliate* to a fully rational credence. Instead, the argument at best shows that an irrational agent who is facing a disagreeing peer has the option of turning rational before conciliating to a fully rational credence. However, this option is nothing new. The irrational agent always had the option of repairing her irrational credence, and then conciliating from her revised (rational) position to a fully rational credence. The General Rationality argument could, at most, establish a view on which conciliating is required of agents only once they are at a rational credence, making it a version of Restricted Conciliationism. But there is a problem with Restricted Conciliationism.

3.1 Resisting the Restricting Response

To get an initial feel for the problem with Restricted Conciliationism, consider a safety requirement to wear a helmet when riding a bike. Presumably, the only people who benefit from wearing a helmet are those who will suffer an accident. Since we do not know in advance who will suffer an accident, we think that everyone should wear one. Suppose someone suggested that the requirement to wear a helmet applied only to agents who were *not* going to suffer an accident. Such a suggestion would be perverse. It would get the motivation for wearing a helmet precisely backwards. If we assume that those who will be in an accident need not wear a helmet, it would be hard to explain why anyone else should. But since this conclusion is absurd (surely some of us should be wearing helmets)

we should reject the suggestion. Analogously, if there is any motivation for conciliating with our disagreeing peers, it comes from those cases where we are the ones who are irrational. But Restricted Conciliationism, oddly, requires conciliation only of initially rational agents.

The analogy above suggests that Conciliationism goes awry when it excludes the initially irrational from having to conciliate. The thought should come naturally to those who share David Christensen's (2007) view regarding the good news about the conciliatory strategy:

> The fact of disagreement is old, but bad, news; it is bad because it indicates the relatively benighted conditions under which we work. But adjusting our beliefs in the direction of those peers with whom we disagree should be welcomed as a valuable strategy for coping with our known infirmities. (Christensen 2007, p. 216)

If only the initially rational must conciliate, the news Conciliationism brings is not good at all. Rather than help us cope with our known infirmities, Restricted Conciliationism stays silent about what to do when those infirmities lead us astray, while ensuring that our peers' infirmities lead us astray too. If Conciliationism is to be any kind of good news, it cannot reserve the need to conciliate exclusively for the initially rational. The motivation for such a view is hopelessly missing. Call this accusation *No Motivation*.

What makes No Motivation stick is not the mere fact that Restricted Conciliationism

asks us to move away from our initial rational positions on the only occasions it applies. Being initially rational does not automatically shield one from having to revise one's belief when new information comes in. What robs the rule of motivation is the irrationality of the disagreeing interlocutor on the only occasions the rule applies. This point relies on the claim that in a peer disagreement, if one peer is rational, the other must be irrational:

> *Someone's Irrational:* When two peers disagree, at least one of them is irrational.

I defend Someone's Irrational soon. But first, note how different Restricted Conciliationism would be from other restricted rules if it indeed required us to conciliate only with irrational others.[7] Suppose, for instance, that we are wondering whether something like Modus Ponens could be a rule of rationality. Nobody should suggest that if you irrationally believe P and also irrationally believe that if P then Q, you should believe Q.[8] So it makes sense to restrict Modus Ponens to those who rationally believe P and if P then Q. A restricted version of Modus Ponens would be unproblematic, as restricting the requirement to initially rational agents entails nothing worrisome about the recommended doxastic addition. We could wholeheartedly endorse such a requirement, and accept its recommendation when we know it applies to us. Not so with Restricted Conciliationism.

---

7 Thanks to Juan Comesaña for pointing this out.

8 See Harman (1986).

Restricting the need to conciliate to initially rational agents means conciliation is required only when our interlocutors have erred. Admittedly, we would not know in a given disagreement whether the conciliatory requirement applies to us, and so we would not know when our interlocutors are the ones irrational. But we would know that conciliation is required only when an irrational peer's view prompts it. No rational agent should aim to conciliate only with irrational others. Thus, reflectively endorsing Restricted Conciliationism is irrational.

Notice that the fact that we may not reflectively endorse Restricted Conciliationism does not depend on some contingent fact about our abilities to follow it. In this regard Restricted Conciliationism is different from, for example, utilitarian moral rules, which are hardly threatened by potentially lacking our endorsement. It may well be that we ought not endorse a utilitarian moral rule if doing so would undermine our ability to follow it. But that would not show such a rule to be incorrect. In contrast, reflecting on Restricted Conciliationism does not undermine our ability to follow it. Recognizing that the rule requires conciliating only with irrational others robs us of any reason to abide by it even if we can. So there is a principled basis for our not endorsing the rule. The fact that there is a principled reason that no reflective agent should aim to follow Restricted Conciliationism is indicative of the rule's implausibility.

Additionally, while it is tempting to think that our ignorance of which party is the irrational one might help with motivating Restricted Conciliationism, it cannot. Our ignorance of who is irrational can only help motivate a conciliatory strategy that is

indifferent to which peer is irrational. It is uncontroversial that if we knew (for certain) we were the irrational party, we would have to concede, and if we knew that we were the rational party, we would have to stick to our guns. Ignorance of who is irrational can, at best, motivate the thought that *whenever* we do not know who is irrational, we should conciliate just in case. Yet this is not what Restricted Conciliationism demands. Proponents of Restricted Conciliationism would owe a story about why our ignorance of who is irrational warrants conciliating only when we are the rational ones.

The observation that if we knew that we were the rational party we would not conciliate reveals another peculiar implication of Restricted Conciliationism. The rule turns out to apply only while we are unaware that it does. This strikes me as an unwelcome result. I am sympathetic to the more general principle according to which genuine requirements do not cease to apply when we learn we are in a situation in which they apply. Learning we are in a situation where some rule applies usually bolsters the motivation for following it, if anything.

But perhaps rules regarding the rational response to unusual kinds of evidence—like peer disagreement and other information suggestive of our irrationality—should not be expected to behave like other rules. For example, in Williamson's unmarked clock case, it may seem prima facie plausible to say that if the clock reads 12:17 we ought to equally believe it is 12:16, 12:17 and 12:18.[9] This too, however, would be a requirement that

[9] Thanks to an annonymous referee for this point. See Christensen (2010) and Elga (2013) for discussions of the case.

applies only when we do not know it does. If we know the clock reads 12:17, we should believe that and that alone. But thinking about the alleged rule reveals that it is plausibly inaccurate. We may reasonably think we ought to equally believe that a clock reads 12:16, 12:17 and 12:18 if it *seems as though* it points to 12:17. Knowing that the clock seems to point to 12:17 would not undermine our motivation for following this requirement. A clock can seem to show 12:17 even when it is not 12:17—if the lighting conditions are poor, if it is far away etc. So the thought that genuine normative rules allow us to know when we are in the situation in which they apply carries weight.

Now, why think that Someone's Irrational is true? Paradigm examples used to promote conciliatory views, like Christensen's (2007) restaurant case, are ones where we seem to know that someone made an error in reasoning. So it would be an interesting result of No Motivation if it forced conciliationists to reject Someone's Irrational. The argument shows that on Restricted Conciliationism, a feature of disagreement that seems to motivate the view—namely, the peers' belief that one of them is irrational—turns out to undermine it. This is already an unexpected result.

Someone's Irrational also follows if we accept a conception of 'peer' according to which peers share their evidence, and also accept the Uniqueness Thesis:

> *Uniqueness:* Given one's total evidence, there is a unique rational doxastic attitude that one can take to any proposition.[10]

[10] This is White's (2005) formulation of Uniqueness.

Indeed, conciliationists are a prime audience for accepting Uniqueness. Kelly (2010) has argued that Conciliationism is committed to Uniqueness.[11] But even if Conciliationism is not committed to Uniqueness, it is more plausible if the thesis is true. The thought of disagreement as putting pressure on agents to move away from their initial credences is more compelling when a disagreement entails that one of the agents has made a rational error.

Proponents of Restricted Conciliationism could reject Uniqueness, and while that would be enough to resist Someone's Irrational, it would not by itself be enough to resist No Motivation. If Restricted Conciliationism aims to capture what we should do in all possible disagreements, it must be that No Motivation never undermines its recommendation. Yet, whenever we would justifiably believe that one of the disagreeing peers is irrational, we could use No Motivation as basis for dismissing recommendations made by Restricted Conciliationism. So the kind of rejection of Uniqueness required to fend off No Motivation would have to deny that we are ever justified in believing one of the disagreeing parties is irrational. Of course, disagreements can arise over very many bodies of evidence, and the opinions involved can vary to a great extent. If we are never justified in believing one of the disagreeing parties is irrational, it must be that permissive situations are both widespread and highly slack.[12] So the brand of Permissivism required for blocking No Motivation, if we wish Restricted Conciliationism to apply to all disagreements, is quite

[11] See Kelly (2010), pp. 120-121. Ballantyne & Coffman (2012) discuss the matter in depth.

[12] By 'permissive situations' I mean a situation in which the evidence permits more than one maximally rational belief. By 'Permissivism' I mean the denial of Uniqueness. See White (2005).

a radical one. It must maintain that very many bodies of evidence are permissive, and that they allow for widely divergent doxastic states.

Another way to reject Someone's Irrational is to deny that peers share all of their relevant evidence. This would allow both disagreeing peers to be rational in cases where their respective evidence differs. But this rejection of Someone's Irrational would not undermine No Motivation in cases where the disagreement persists even after the agents explicitly share their evidence. So No Motivation is immune to this objection in cases of *full-disclosure*,[13] where disagreeing agents are stipulated to be aware of each other's evidence.

This defense of Someone's Irrational works best on the assumption that an agent's evidence is entirely sharable. But even if our evidence is not always entirely sharable, on many occasions it is. If Restricted Conciliationism is to apply as a rule for belief revision in all disagreement, it must apply when we know we indeed share all our evidence with our peers.

It is possible to resist No Motivation by further restricting Conciliationism, so as to guarantee absence of justification for thinking some party in the disagreement is irrational. Conciliating may thus be taken as required only of initially rational agents who lack justification for believing one of the disagreeing parties is irrational. Such a rule would not only be ad hoc, but also largely incomplete. It is hard enough to grant that a rule regarding the correct response to disagreement can leave out all initially irrational agents, but at least that restriction has a principled reason behind it. As Christensen (2011) suggests, there are

[13] Feldman (2006) introduces this terminology.

many ways to be irrational, and so expecting one rule to account for all possible mistakes may be expecting too much.[14] I see no reason to allow a similar restriction in the case of agents who know they are in a disagreement where someone is irrational. Doing so would seem to do nothing more than limit the application of Restricted Conciliationism for no other reason than to strategically avoid No Motivation.

Returning to the General Rationality argument, it may seem that the kind of Permissivism that it allows fits the bill. Since on Cohen's suggestion generally rational agents get to upgrade their irrational beliefs, situations where peers could justifiably believe that one of them is irrational would be harder to come by. Nonetheless, if the disagreeing parties justifiably believe that neither of them went through Cohen's upgrading procedure, they could justifiably believe one of them is irrational, and use No Motivation to refuse to conciliate. So even a highly permissive conciliatory view like Cohen's provides incorrect and rationally dismissible advice in cases where the peers know one of them is irrational. Here too, restricting the view so as to leave these cases out would be ad hoc.

In what follows I show that other possible responses to EB are unsuccessful. Some fall prey to No Motivation. Others do not tell agents what to believe in cases of disagreement. Others still sanction widespread rational dilemmas.

4 The Necessary Requirement Response

An attractively simple way around EB involves construing Conciliationism as only a

[14] Christensen (2011), p. 4.

necessary requirement of rationality. EB does not gain traction unless we treat conciliation as sufficient for attaining a rational belief. Kelly (2014) calls the construal of Conciliationism that exposes it to EB *Strong Conciliationism*:

> *Strong Conciliationism:* An agent's holding the conciliated opinion is both sufficient as well as necessary for her opinion to be fully reasonable or justified.

On strong conciliatory views, conciliating with a disagreeing peer is required of all agents, and doing so suffices for reaching rational beliefs. If instead we interpret Conciliationism as imposing just a necessary requirement, conciliating would not guarantee a rational belief. Kelly names such an interpretation *Weak Conciliationism*:

> *Weak Conciliationism:* A perfectly rational agent would not hold some opinion that is out of step with the conciliated opinion.

Weak Conciliationism implies that conciliating with a disagreeing peer is necessary for holding a rational credence. Since it is not committed to conciliation being sufficient for holding a rational credence, it does not allow easy bootstrapping.

Notice that as stated, Weak Conciliationism tells us little about how agents should respond to peer disagreement. But precisely what we want to know is what is required of

any agent in a peer disagreement situation.[15] This is not to say that the direction taken by Weak Conciliationism is hopeless. Perhaps some EB-avoiding requirement can be found nearby. So to give Weak Conciliationism due consideration, we must look at possible requirements in the same spirit.

One natural way of turning Weak Conciliationism into a requirement that tells us how to respond to disagreement is this:

> *Necessary Conciliationism:* If one believes P to degree d_1, and one encounters a peer believing P to a different degree d_2, then it is necessary for having a rational belief that one conciliate to an intermediate degree d_3.

According to Necessary Conciliationism, rationality requires that all agents conciliate upon meeting a disagreeing peer. Since it is only a necessary requirement, meeting it does not guarantee that one ends up with a rational belief. In particular, if the rule is to avoid EB, those who are initially irrational would remain irrational even if they conciliate.

However, the conciliatory move that Necessary Conciliationism deems necessary is sometimes entirely unnecessary. Suppose that agent S_1 starts out by misjudging her evidence E, and assigns an unsupported credence d_1 to P. In the meantime, S_1's peer, S_2,

---

[15] For example, Elga (2007) asks: "How should you take into account the opinions of an advisor?" Enoch (2010) asks: "How should you update your (degrees of) belief about a proposition when you find out that someone else—as reliable as you are in these matters—disagrees with you about its truth value?" Other authors including Christensen (2009) present the question similarly.

also evaluates E poorly and assigns a different unsupported credence d_2 to P. Once the two agents meet, S_1's total evidence supports some credence d_4 other than the one (d_3) required by Necessary Conciliationism—for if it supported d_3 then conciliating would suffice for a fully rational belief. At that point, S_1 could (and should) realize that all of her evidence justifies d_4 and move directly to that credence. At no point would it be necessary for S_1 to assign d_3 to P, contra Necessary Conciliationism.

Two other interpretations of Weak Conciliationism that may serve as responses to EB can be found in the dialectic. I discuss those next.

4.1 The Wide Scope Response

Christensen (2011) offers a response to EB that seems to fit well with Weak Conciliationism:

> Conciliationism tells us what the proper response is to one particular kind of evidence. [...] But having taken correct account of one bit of evidence cannot be equivalent to having beliefs that are (even propositionally) rational, all things considered. If one starts out by botching things epistemically, and then takes correct account of one bit of evidence, it's unlikely that one will end up with fully rational beliefs. And it would surely be asking too much of a principle describing the correct response to peer disagreement to demand that it include a complete recipe for undoing every epistemic mistake one might be making in one's thinking. (Christensen 2011, p. 4)

He goes on to clarify:

> Conciliationism does not entail that Right and Wrong [the disagreeing agents of whom one is rational and the other irrational] end up with equally rational beliefs [after conciliating]. Nor does it entail that they were rationally mandated to make equally extensive revisions to their opinion. Of course, it does have the consequence that the revisions called for by the disagreement are equally extensive. (Christensen 2011, pp. 4-5, my brackets)

If Christensen is right and conciliatory views are silent about what credences agents ought to hold, then EB appears to be misguided. On this view, Conciliationism is merely a view about the proper response to one piece of evidence, or the amount of revision called for by the disagreement alone, and not a view about how to find an all around rational belief on any disputed matter.

As with Weak Conciliationism, it is not clear what conciliatory requirement follows when we take conciliating to merely be the correct response to one bit of evidence. Cohen (2013) suggests that Christensen has a wide scope requirement in mind:

> *Wide Scope Conciliationism:* One ought see to it that if one believes P to any degree different than that to which a peer does, then one conciliates.

Having the entire conditional in the scope of the normative operator allows for two ways of satisfying the requirement once a disagreement arises. Wide Scope Conciliationism will be satisfied if either its antecedent is made false or its consequent is made true. Since Wide Scope Conciliationism is silent about which of the two ways of satisfying it is required in which situations, it is able to avoid EB by not requiring the irrational to conciliate. Of course, this does not mean that any way of satisfying Wide Scope Conciliationism would be perfectly rational. If conciliating were always available as a perfectly rational way of satisfying it, EB would return. Being silent about which of the two ways of satisfying Wide Scope Conciliationism is ultimately required makes the rule compatible with the thought that in certain situations other rational rules will forbid conciliating.

The wide scope reading of Conciliationism appears to fit Christensen's description of agents who conciliate as taking "correct account of one bit of evidence." If an initially irrational agent conciliates in the face of disagreement, then by doing so she successfully avoids violating a wide scope requirement. Moreover, because two ways of complying with Wide Scope Conciliationism are available, one may comply with it and still fail to have a rational belief. So this interpretation of Christensen is faithful to Weak Conciliationism.

Looking a bit further into this interpretation, however, reveals a problem. As it stands, Wide Scope Conciliationism still does not tell us how agents should respond to peer disagreement. It does not tell us which agents should satisfy the consequent and conciliate, and which should do something else. Plausibly, when the agent is the irrational party, the only rational way for her to satisfy Wide Scope Conciliationism would be to do something

other than conciliate. Indeed, this must be true for the principle to avoid EB. But this also leaves rational agents as the only ones who should satisfy the principle by conciliating. So even if Wide Scope Conciliationism is correct, it remains true that only the initially rational must conciliate. This is all that is needed for No Motivation to apply here as well. It therefore turns out that a reasonable understanding of Wide Scope Conciliationism has the implication that only initially rational agents must conciliate when facing a disagreeing peer.

4.2 The Special Requirement Response

A different response to EB that is friendly to Weak Conciliationism can be derived from Miriam Schoenfield's discussion of Calibrationism:

> We should not think of calibrationism as a principle of epistemic rationality, where such principles are understood as principles about the relationship between bodies of evidence and belief states. Rather, we should think of calibrationism as a principle of reasoning. I am understanding principles of reasoning as principles about which transitions of thought, or reasoning processes, should occur in the process of deliberation. (Schoenfield 2014, pp. 452-3)

If we adapt Schoenfield's idea to Conciliationism, we get a view according to which conciliating is required by the rules of *correct reasoning*. We can capture the point by revising Natural Conciliationism:

> *Reasoning Conciliationism*: If one believes P to degree d_1, and one encounters a peer believing P to a different degree d_2, then correct reasoning requires one to conciliate to an intermediate degree d_3.

According to Reasoning Conciliationism, conciliating in a peer disagreement is a requirement of correct reasoning rather than a requirement of rationality. This means that in cases of disagreement, conciliating is required even if the resulting credence would be irrational. Thus, while conciliating is necessary, a rational credence is not guaranteed for those who conciliate, and EB is avoided.

Though Schoenfield's suggestion arises in a different context, it provides an appealing response to EB. While rationality may require us to revise our beliefs one way, correct reasoning may require us to revise another way, and no conflict between rational requirements would arise. The basic move is to distinguish between different kinds of requirements, or different senses of 'ought,' and say that Conciliationism imposes a special kind of requirement, or that it employs a special kind of 'ought.'[16]

Despite its initial appeal, we should reject Reasoning Conciliationism and suggestions like it. One concern is that such suggestions multiply kinds of normative requirements (or senses of 'ought') beyond necessity. But even if we think that this multiplication is called for, a more significant worry looms. It is not clear how reasoning requirements are related to rationality. If correct reasoning is not necessary for rational belief, then agents can

[16] Worsnip (2018) defends a similar idea.

rationally ignore reasoning requirements. Alternatively, suppose correct reasoning is indeed necessary for rational belief. Reasoning Conciliationism would then be equivalent to Necessary Conciliationism, sharing its implausible demands. The same can be said of other interpretations of Conciliationism that take the normative operator it employs to be anything other than the 'ought' of rationality.

4.3 The Dilemma Response

When discussing disagreement about disagreement, Christensen (2013) concludes that in certain situations epistemic requirements may conflict:

> …the fact that there is this tension among our epistemic ideals need not mean that any of them is incorrect. It might just mean that in certain situations (in particular, when one gets good evidence against the correctness of what are in fact the correct ideals), one will end up violating some ideal or other, no matter what one ends up believing. (Christensen 2013, p. 91)

This line could be turned into a response to EB: when an irrational agent encounters disagreement, she faces a genuine epistemic dilemma.[17] Call this the *Dilemma Response*.[18]

Consider agents S_1 and S_2 from the example in section 4. S_1 has an irrational credence

[17] Thanks to Brian Weatherson for this observation.

[18] This line can also be used to respond to my criticism of Necessary Conciliationism and Reasoning Conciliationism.

d_1, while S_2 has a different irrational credence d_2. After meeting each other, S_1's total evidence supports the adoption of some credence d_4 different from the one (d_3) ordered by Conciliationism (to ensure avoiding EB). According to the Dilemma Response, since S_1 is rationally required to conciliate to d_3, she faces a conflict between what her total evidence requires and what Conciliationism requires. No matter how S_1 revises, she would violate some rational requirement. S_1 would be in a genuine rational dilemma.

The view promoted by the Dilemma Response should be understood as one on which the conflicting rational requirements are necessary rather than sufficient for attaining rational beliefs. If both the requirements were sufficient, then following either would suffice for reaching a fully rational belief, making such a view subject to EB. Alternatively, considering one of the requirements sufficient and the other necessary would be incoherent. If by following the sufficient requirement one can reach a fully rational belief then following the necessary requirement would not truly be necessary. But if both requirements are necessary for having a fully rational belief, then the dilemma is serious. No matter what the agent does, she would fall short of doing all that is required for reaching a fully rational belief.[19]

The Dilemma Response to EB is a kind of bullet-biting response. It successfully avoids EB, since an initially irrational agent would violate some requirement no matter what she does, and no easy path to rational belief would be available to her. But recall that EB is an

[19] Note that if there is a rational way out of an epistemic dilemma which would land the agent in a fully rational belief state then the dilemma was not a genuine one all along.

instance of the detaching problem, and can be understood as accusing Conciliationism of giving rise to conflicting rational requirements in cases where an irrational agent encounters disagreement. So on the dilemma response we have to accept much of what the detaching problem targets as problematic.

Yet there is a more significant problem with the Dilemma Response than its allowing for genuine rational dilemmas to arise. If Christensen is right in the above quote, rational dilemmas are already part of epistemic life. The more significant problem is that rational dilemmas would arise all too frequently. The detaching problem comes up frequently, since we adopt irrational beliefs frequently. Biting the bullet in response to an instance of it (in the case of EB) would seem to commit the biter to doing the same when the problem comes up elsewhere, thereby granting that rational dilemmas are widespread. This would be a high theoretical cost. A view that admits of frequent rational dilemma fairs worse as a view of what we ought to believe than one that does not. In the absence of reason to think that the case of Conciliationism is special and warrants a unique response to the detaching problem, biting the bullet risks sanctioning widespread rational dilemmas.

5 Easy Bootstrapping and Total Evidence

The disagreement debate has recently led some participants to provide accounts that explicitly aim for greater generality than a mere necessary condition on rational belief given disagreement. For example, here are two such suggestions:

Christensen's (forthcoming):

> *Idealized Thermometer Model:* The credence in C it would be rational for the agent to form, given *all* her evidence, is the credence in C that would be rational independent of C's support from first-order considerations, conditional on a relevantly similar agent adopting credence *n* in C on the basis of first-order support from the agent's evidence.[20]

Paulina Sliwa and Sophie Horowitz's (2015):

> *Evidential Calibration:* When one's evidence favors P over ~P, one's credence in P should equal the expected reliability of one's educated guess that P.

Rules such as these describe how we should integrate information about our *expected reliability*[21] with our other evidence.[22] Since disagreement plausibly provides information regarding our expected reliability, the recommendations issued by such rules should include the rational response to disagreement.

Rules that aim to take all of our evidence into account clearly have potential for avoiding the Easy Bootstrapping argument. The directions these rules provide are unlikely

---

[20] Here Christensen assumes *n* to be the rational credence to have given one's first-order considerations. He also assumes takes 'relevantly similar' to mean "someone who is similar with respect to the reliability evidence the agent has about herself."

[21] Sliwa and Horowitz (2015) define 'expected reliability' as one's expected propensity to guess correctly.

[22] More generally, they aim to tell us how to weigh our first-order and higher-order evidence. See Kelly (2010) for discussion of higher-order evidence.

to license two irrational agents to merely move toward one another and call it a day.[23] However, this way of overcoming EB comes at the price of abandoning the kind of Conciliationism that has been the target of discussion all along. The kind of Conciliationism discussed so far involves the thought that the rational response to disagreement is a move toward one's peer, or at least a move in the direction of what our peer takes to be more likely between P and not-P.[24] Truly Conciliatory views cannot allow, for example, that sometimes when we start at a low credence and meet a peer at an even lower credence we must move to a much higher credence.

If total evidence rules are to avoid EB, they will allow such cases. Suppose that agents S_3 and S_4's shared evidence E strongly favors P. Assume that before meeting each other, both irrationally take E to favor not-P, with S_3 assigning a credence of .2 to P, and S_4 assigning .1. Then they meet. In order to avoid EB, a total evidence rule must not recommend that the agents merely settle on some close-by credence under .2. But if a rule recommends that the agents' credences be much closer to what the shared evidence in fact supports—perhaps

---

[23] Christensen specifically mentions Kelly's EB worry immediately after offering his model, saying that the model "puts rational credences" where traditional Conciliationism "had the agent's actual initial credences." Here Christensen is talking about what he calls the Standard Thermometer Model as the relevant conciliatory view.

[24] For example, Elga (2010) says that on conciliatory views, "finding out that a respected advisor disagrees with one should move one at least a little in the direction of the advisor's view." The idea that respecting our peers involves moving toward their formed credences also seems to be the one behind the difference-splitting recommendation of the equal weight view. See Kelly (2010) and Cohen (2013). Christensen (2009) notes that there may be certain exceptions, where two high credences warrant a move to an even higher one. See Easwaran et al. (2016) for a development of such a view.

.5 or even higher—it would not be conciliatory in any plausible sense of the term. What unifies conciliatory views is their taking our peer's view as a pretty strong indication of what our own view should be. It would be a slap in the face of my peer if upon learning that she is at .1 when I am at .2 I move to .5.

Now it may still be that in so revising I would be giving my peer's opinion equal weight to my own, if I give both little weight. Doing that would make the move from .2 to .5 technically consistent with an understanding of Conciliationism on which all it takes to conciliate is giving a peer's opinion equal weight to one's own.[25] But technically consistent is not consistent enough. It is technically consistent with the equal weight view to give no weight to a peer's opinion, if one gives her own opinion no weight either. Yet such a weighing of opinions would allow complete dismissal of a peer's opinion. Surely this is not what conciliationists have in mind.

Having said that, it is still possible to start at a low credence, meet a peer at an even lower one, and end up with a much higher credence while also giving the peer's opinion significant weight. This can be done if we reassess the shared evidence and come to a different judgment than our initial low one. For example, S_3 could come to a new judgment that the evidence supports a .9 credence in P instead of her initial .2, and only then take her peer's .1 into account. However, any total evidence rule that says this is what

[25] There is already a wrinkle in this understanding of the equal weight view. It is not clear what it means to give both one's own judgment and a peer's judgment little weight, if one ultimately ends up giving all the weight to her own later assessment of what the originally shared evidence supports.

agents ought to do opens itself up to a No Motivation kind of argument.[26] If EB is to be avoided then it must be that the reassessment of the evidence is done rationally. But if that is done rationally, and we form the belief the shared evidence supports, conciliating would again turn out to only be required of agents only once they have correctly assessed the shared evidence.

There is, therefore, a tradeoff between how truly conciliatory total evidence rules can be and their success in avoiding the bootstrapping problem. The more conciliatory a rule is, the more weight it must give to the peers' initial credences, and the closer it comes to allowing easy bootstrapping.

6 The upshot for Conciliationism

The Easy Bootstrapping argument shows that conciliating is not what everyone needs to do to reach a rational belief upon meeting a disagreeing peer. Since there are many ways to arrive at an irrational belief, it may be too much to expect a concise rational rule about disagreement to guide all agents in undoing their previous mistakes.[27] This observation, however, is not grounds for dismissing the bootstrapping concern. The challenge for conciliationists is to formulate a plausible requirement that does not allow bootstrapping. I

[26] Christensen (forthcoming) explicitly says that the Idealized Thermometer Model does not require such a step-by-step procedure.

[27] Some doxastic revision rules can still do this due to the strength of evidence involved. For example, an infallible oracle's testimony regarding the truth of P arguably suffices for reaching a rational belief about P merely by adopting the oracle's own view.

have explored two kinds of ways conciliationists attempt to meet this challenge, namely, restricting Conciliationism only to rational agents, and treating Conciliationism as necessary but not sufficient for fully rational beliefs.

When we restrict Conciliationism, we are left with a perverse rational requirement. If we require only those who responded to their evidence rationally to conciliate, then the view is hopelessly unmotivated. When we treat Conciliationism as a necessary requirement, we are left wondering how an agent should revise her beliefs in response to a disagreeing peer. Since this is the central question in the disagreement debate, a view that says nothing more than that is ultimately unsatisfying. I have looked into a few ways of developing this response: Necessary Conciliationism, Wide Scope Conciliationism, and Reasoning Conciliationism. These views either collapse into Restricted Conciliationism, or deem necessary an unnecessary doxastic revision.

A further move available to conciliationists is to bite the bullet and admit that Conciliationism sometimes leads to genuine conflicts between rational requirements. I argued that since EB is an instance of the pervasive detaching problem, this strategy would suggest that such dilemmas arise when the detaching problem does, which is too often.

Finally, I mentioned total evidence rules as potentially viable for telling us how a disagreeing peer should affect our beliefs. I argued that if such rules are to avoid a bootstrapping problem, they cease to be conciliatory in spirit, as they would give too little weight (even if equal weight) to our disagreeing peers. This is not to say that no total evidence rule can be right, of course. It is just to say that it will not save the intuitively

compelling kind of Conciliationism, which takes a peer's dissenting view to be a strong indication of what the shared evidence supports.

Works cited

Ballantyne, Nathan. & Coffman, E.J. (2012). "Conciliationism and Uniqueness". *Australasian Journal of Philosophy* 90: 657-670.

Broome, John. (1999). "Normative Requirements". *Ratio* 12: 398–419.

Christensen, David. (2007). "Epistemology and Disagreement: The Good News". *Philosophical Review* 116: 187–217.

–––– (2009). "Disagreement as Evidence: The Epistemology of Controversy". *Philosophy Compass* 4(5): 756-767.

–––– (2010). "Rational reflection". *Philosophical Perspectives* 24: 121–140.

–––– (2011). "Disagreement, Question Begging, and Epistemic Self Criticism". *Philosophers Imprint* 11(6): 1-22.

–––– (2013). "Epistemic Modesty Defended". in Christensen and Lackey (eds.), *Disagreement: New Essays* (Oxford: Oxford University Press), 77-97.

–––– (Forthcoming). "Disagreement, Drugs etc.: From Accuracy to Akrasia". *Episteme.*

Cohen, Stewart. (2013). "A Defense of the (almost) Equal Weight View". In Jennifer Lackey and David Christensen (eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 98–120.

Elga, Adam. (2007). "Reflection and Disagreement". *Nous* 41(3): 478–502.

–––– (2010). "How to disagree about how to disagree. In R. Feldman and T. A. Warfield (Eds.), Disagreement, Oxford University Press. 175–186.

–––– (2013). "The puzzle of the unmarked clock and the new rational reflection principle". *Philosophical Studies* 164: 127–139.

Enoch, David. (2010). "Not Just a Truthometer: Taking Oneself Seriously (But Not Too Seriously) in Cases of Peer Disagreement". *Mind* 119 (476): 953–997.

Easwaran, Kenny et al. (2016). "Updating on Credences of Others: Disagreements, Agreement and Synergy". *Philosophers' Imprint* 16(11): 1-39.

Feldman, Richard. (2006). "Epistemological Puzzles About Disagreement" in Stephen Hetherington (edd.) *Epistemology Futures* (Oxford University Press, Oxford), 216 -236.

–––– (2007). "Reasonable Religious Disagreement". In L. Antony (ed.), *Philosophers Without Gods*. Oxford University Press, 194-214.

Finlay, Stephen. (2010). "What *ought* probably Means, and why you can't detach it". *Synthese* 177: 67–89.

Harman, Gilbert. (1986). Change in View: Principles of Reasoning, MIT Press, Cambridge, MA.

Kelly, Thomas. (2010). "Peer Disagreement and Higher-Order Evidence". In T. Warfield and R. Feldman (ed.), *Disagreement*. Oxford University Press, 111–75.

–––– (2014). "Believers as Thermometers". In Jonathan Matheson and Rico Vitz (eds.), *The Ethics of Belief*, Oxford: Oxford University Press, 301–314.

Kolodny, Niko. & MacFarlane, John. (2010). "Ifs and Oughts". *The Journal of Philosophy* 107(3): 115-143.

Schoenfield, Miriam. (2015). "A Dilemma for Calibrationism". *Philosophy and Phenomenological Research* 91(2): 425-455.

Sliwa, Paulina. & Horowitz, Sophie. (2015). "Respecting All the Evidence'. *Philosophical Studies* 172(11): 2835-2858.

Weatherson, Brian. (ms.). "Do Judgments Screen Evidence?" available at <http://brian.weatherson.org/JSE.pdf.>

White, Roger. (2005). "Epistemic Permissiveness". In J. Hawthorne (ed.), *Philosophical Perspectives* 19, Epistemology, Malden, MA: Blackwell Publishing, 445–59.

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

Normative Detaching Undermines Principles of Higher-Order Evidence

Eyal Tal

University of Arizona

Abstract: Certain attitudes come with strings attached. For instance, believing that chocolate is harmful to dogs commits us to not feeding our dogs chocolate. Focusing on doxastic attitudes, epistemologists habitually offer principles that aim to capture what beliefs we are committed to adopting or avoiding given other beliefs that we have. A known worry with some of these principles is that they seem to open the door to troublesome rational dilemmas. If believing ~P commits us to disbelieving P, and we believe ~P despite evidence to the contrary, then our commitment would have us disbelieve something that our evidence requires we believe. Attempts to avoid the problem share the result that only a rationally held belief commits its holder to adopting or avoiding other beliefs. This result has a surprising implication for possible principles regarding the rational response to higher-order evidence. If higher-order evidence requires belief revision only of agents who responded rationally to their original evidence, then higher-order evidence requires belief revision only when it is misleading. It would be perfectly rational of agents who notice this never to adhere to those principles, knowing that they do nothing but mislead. As a result, a variety of plausible-sounding principles regarding the rational response to higher-order evidence are incorrect.

1 Introduction

Certain attitudes carry commitments. Such attitudes make it so that agents who hold them may have to adopt further attitudes, or perform some actions, that other agents do not have to. For instance, thinking that smoking is harmful commits you to thinking it is not

healthy. Despising your smoking habit commits you to intending to quit. Any agent who violates commitments that her attitudes carry would not be as she ought to be, in some relevant respect, due to the violation. An agent who believes that smoking is both harmful and healthy is incoherent, and arguably irrational.[1] An agent who despises her smoking habit but does not intend to quit is akratic.

It would be valuable to find principles that capture which attitudes commit us to what, and in which circumstances. Here are some attractive such principles:

- If you believe P you should disbelieve ~P.
- If you aim to φ and know that a necessary mean to φ is ψ, you should aim to ψ.
- If you love S and you know that S needs help, you should attempt to help S.

Hypothetical imperatives like these share a structure:

> *Commitment*: If you hold attitude A in situation B, you ought to adopt attitude (or perform action) C.

Examples of Commitment principles are prevalent in philosophical debate. A naïve rendition of van Fraassen's Reflection Principle is one example of such a principle:

---

[1] See Worsnip (2018) for a distinction between coherence requirements and rational requirements.

> *Reflection:* If you believe that you will rationally believe P in the future, you ought to believe P now.

At first glance, Reflection may seem right. There is no excuse for avoiding a view when you know that evidence for it is coming your way. But while Reflection may seem right, it has a weakness. Consider a case where you should not believe that your future self rationally believes P—perhaps because you have no evidence regarding the beliefs of your future self. Unfortunately, Reflection implies that even if you irrationally believe that your future self rationally believes P, you should still believe P now.

The weakness we find in Reflection pervades many principles that share the structure of Commitment. This weakness is none other than the *Detaching Problem*, to which many responses have been offered. I argue that responses to this problem imply that only initially *righteous* agents must adopt the attitudes (or perform the actions) that their other attitudes commit them to. That is to say, only agents who *should be* holding attitude A to begin with, ought to adopt attitude (or perform action) C. But this implication spells trouble for possible principles regarding the required response to higher-order evidence—evidence about what we should believe. Principles that require only already rational agents to revise their beliefs due to higher-order evidence are perverse. Rational agents, by definition, responded rationally to their initial evidence. So following these principles is guaranteed to draw those agents away from the rational belief that they hold given their original evidence, due to nothing more than misleading information about their rationality. Noticing this

makes it rational to reject any principle that requires only the initially rational to revise their beliefs due to higher-order evidence. It follows that the prospects for finding principles regarding the rational response to higher-order evidence are bleak.

2 The Detaching Problem

The weakness had by Reflection flows directly from a more general problem, known as the Detaching Problem. Detaching problems infect many hypothetical imperatives that share the structure of Commitment. Following Stephen Finlay's presentation of the problem, consider the principle of self-reliance:

> *Self-Reliance:* If an agent S believes that she ought to φ, then she ought to φ.

Finlay says Self-Reliance gives rise to an apparent contradiction:

> There is something intuitively right about this principle, which captures the platitude that one ought to follow one's conscience. If Jorja believes that she ought to skip school and yet fails to (intend to) do so, then she or her behaviour is in some way defective (*irrational*, some say). But there also seems to be something wrong with the principle. It appears to imply something implausible: that a person can never be mistaken about what she ought to do. By modus ponens, if a conditional and its antecedent are both true, we can *detach* the consequent. From the Self Reliance Principle and Jorja's believing she ought to skip school, we

should therefore conclude simply that Jorja ought to skip school. (Finlay 2010, p. 68)

Let us distinguish a *good* case from a *bad* case regarding Jorja's belief that she should skip school. In the good case, Jorja believes that she ought to skip school for good reasons—perhaps a close relative of hers has just passed away, and in order to attend the funeral she must skip school. Here, Self-Reliance gives rise to no problem. Given that Jorja's belief is perfectly justified, she really ought to skip school. In the bad case, Jorja believes that she ought to skip school for bad reasons—perhaps she thinks that she should live every day as if it were her last, and this involves going to the beach instead of school. Here, Self-Reliance does have an undesirable consequence. Since the antecedent of the principle is true of Jorja and skipping school, it follows that she ought to skip school. But this is odd. Intuitively, having an unjustified belief about having to skip school is not enough to make it so that Jorja indeed ought to skip school. It is much more plausible that Jorja ought to abandon her unjustified belief, and not skip school. Thus, a conflict arises in the bad case between what we intuitively think that Jorja should do and what Self-Reliance tells us that she should.

It is often easy to generate a detaching problem for principles sharing the structure of Commitment. For many such principles, all we would have to do is stipulate a bad case in order to show that they give rise to normative conflicts. In bad cases, agents form the relevant initial attitude when they should not, and we would have the intuition that they

should abandon our initial attitude. This intuition would go against what principles sharing the structure of Commitment require, namely, that agents take the further steps that their held attitudes commit them to.

Reflection and Self-Reliance are clear examples of where a detaching problem comes up. Another, less obvious, example of the detaching problem at work, is Thomas Kelly's (2010, 2014) recent objection to conciliatory views of peer disagreement (Conciliationism).[2] Kelly's idea is this: take an agent who irrationally assigns some credence to P, and introduce her to a peer who irrationally assigns a different credence to P. Then, the argument goes, if Conciliationism is correct, both agents would be rationally required to conciliate with each other. And if such agents are genuinely required to conciliate with each other, their post-conciliation credence must be rational for them to have. So, according to Kelly, when two initially irrational agents conciliate with each other, they arrive at fully rational credences all too easily. Kelly calls this the Easy Bootstrapping objection, after the intuitively illegitimate ease with which agents get to turn their irrational beliefs rational.[3]

Looking carefully at Kelly's objection, however, reveals a close match to the aforementioned recipe for detaching problems. Conciliationism may naturally be understood as sharing the structure of Commitment, roughly along these lines:

---

[2] According to Conciliationism, the rationally required response to a disagreeing peer is to revise one's view in the direction of the peer's, to some considerable extent. See, for instance, Christensen (2007), Elga (2007).

[3] Weatherson (ms.) names this phenomenon 'epistemic laundering'.

> *Conciliationism:* If S believes P to degree d_1 and encounters a peer who believes P to a different degree d_2, then S ought to conciliate to an intermediate degree d_3.

Considering a bad case is all that is needed to get a detaching problem going. An agent who irrationally believes P to degree d_1 would still be required to revise in the direction of her peer's credence, as Conciliationism demands. But this would arguably pave an unacceptably easy path for the agent to expunge her irrational record and form a fully rational credence. It may be that, as with Jorja's desire to skip school, what such an agent should really do is repair her initial belief before anything else.

In what comes next I look into some prominent solutions to the detaching problem, and argue that they share an interesting implication. In order to avoid the detaching problem, all solutions end up denying that agents holding an attitude irrationally ought to follow through on the commitments these attitudes carry. This means that if anyone's attitudes commit them to other ones, it is agents in the good case, i.e., agents whose attitudes are justified. I then argue that it is just those agents who have no reason to let higher-order evidence affect their beliefs.

3 Solutions to the detaching problem

We find a number of potential solutions to the detaching problem in the literature.[4] These

[4] See Silk (2014) for a representative list.

solutions sometimes appear as general responses to the detaching problem, and sometimes appear in the context of defending a specific principle from a detaching problem. Because the problem is a general one, the latter, local solutions easily fit more general categories. In this section I distinguish four kinds of approaches that hope to get around the detaching problem. I restrict discussion to principles regarding the rationality of belief that share the structure of Commitment. So I focus on principles of the following form:

> *Doxastic Commitment:* If you believe P and situation B arises, you ought to believe Q.

As before, I will refer to cases in which we initially believe P rationally as *good* cases, and to those where we initially believe P irrationally as *bad* cases.

I also restrict discussion to attempts of *resolving* the detaching problem. Some may think that adopting an attitude we should not adopt lands us in a kind of inescapable dilemma—so that no matter what we believe or do we would violate some norm.[5] Such concession to the detaching problem has the upshot that normative dilemmas arise too frequently. All it would take to land ourselves in such a dilemma would be for us to adopt some attitude unjustifiably—as we all do all the time. This approach also goes against the intuitive thought that when we adopt an irrational attitude we really should abandon it rather than follow up with what it commits us to. For instance, if S irrationally believes that she should

[5] This kind of response can be seen in Christensen (2013).

assassinate the mayor, she should obviously abandon her belief rather than start planning the assassination. In what follows, then, I discuss four responses to the detaching problem, all of which deny that we face genuine dilemmas in bad cases.

3.1 The commitments apply in good cases only

On what we may call the *Good Case* approach, only justifiably held attitudes commit us to other attitudes or actions.[6] That is, only agents in the good case are required to take the further steps their attitudes commit them to. Applying this approach to Doxastic Commitment yields the following:

> *Good Case Doxastic Commitment:* If you rationally believe P and situation B arises, you ought to believe Q.

The principle only requires agents whose initial beliefs are rational to revise their beliefs according to the relevant commitments. Indeed, this solution quickly comes to mind when considering a defense of Reflection from the detaching problem. Applied to Reflection, the Good Case approach would make it so that only those who rationally believe their future selves to rationally believe P are required to believe P now. In rule form:

---

[6] Weatherson (ms.) and Cohen (2013) suggest defending Conciliationism from Kelly's objection (i.e., from a detaching problem) in this way. Way & Whiting (2015) defend the principle: If you justifiably believe that you ought to φ, you ought to φ.

Good Case Reflection: If you rationally believe that you will rationally believe P in the future, you should believe P now.

A detaching problem only surfaces when a Commitment principle requires an agent in a bad case to take steps that her irrational attitude appear to commit her to. In those cases, intuition has it that the agent should do her best to fix her inadequate attitude instead of taking the relevant further steps. So, by restricting Reflection to agents in the good case only, Good Case Reflection (and the Good Case approach in general) avoids a detaching problem by requiring nothing of initially unjustified agents.

3.2 The normative operators take wide scope

On the *Wide Scope* approach, the correct way to understand Commitment principles is as imposing a wide scope requirement.[7] That is, Commitment principles are correct if taken to mean that their normative operators have the entire conditional in their scope. Applied to Doxastic Commitment, the Wide Scope approach gives us this:

Wide Scope Doxastic Commitment: You ought to see to it that if you believe P in situation B, you believe Q.

Having the normative operator take wide scope allows for two ways of complying with the

[7] See, for instance, Broome (1999), Way (2010).

requirement.[8] We can ensure the truth of the conditional either by satisfying its consequent or by falsifying its antecedent.

The Wide Scope approach provides requirements that, strictly speaking, do not require us to believe anything in particular. Instead, it forbids certain combinations of attitudes in certain situations. As we have seen, only principles that require agents in the bad case to follow up on their apparent commitments face a detaching problem. So by not requiring that, the Wide Scope approach seems able to avoid the problem. Applying this approach to Reflection illustrates how it works:

> *Wide Scope Reflection:* You ought to see to it that if you believe you will rationally believe P in the future, you believe P now.

An agent who rationally believes that her future self rationally believes P could comply with Wide Scope Reflection by believing P. An agent who irrationally believes that her future self rationally believes P could comply with Wide Scope Reflection by abandoning her irrational belief. Having two options for how to comply with Wide Scope Reflection provides all agents, be they in the good or in the bad case, with an intuitively unproblematic route to obeying the principle. And because agents in the bad case are not required by the principle to believe anything in particular, a detaching problem does not come up.

---

[8] Here I leave out a way of complying with the principle by undoing the relevant situation B.

Like the Good Case approach, the Wide Scope approach also has some prima facie appeal. It manages to avoid the detaching problem by never requiring agents in the bad case to form further beliefs that their irrational beliefs appear to commit them to. Instead, what it requires is that the agent not violate a certain prohibition. Since there are multiple ways of avoiding violating the prohibition, the fully rational way to do so may well depend on the relevant circumstances. This is evident in Wide Scope Reflection, where agents who should not believe that their future selves rationally believe P should simply abandon their irrational belief, while agents who should believe that their future selves rationally believe P should go ahead and believe P.

3.3 Some normative operators are special

On the *Special Ought* approach, the 'ought' that appears in Commitment principles has a unique meaning.[9] Instead of the familiar 'ought's of morality or rationality, the 'ought' employed in these principles draws its normative force from a different source—perhaps that of coherence, or reasoning, for example. Applied to Doxastic Commitment, we get:

> *Special Ought Doxastic Commitment:* If you believe P and a situation B arises, you special-ought to believe Q.

[9] Worsnip (2018) defends a view of this sort, and combines the thought with a wide scope interpretation of Commitment principles. Schoenfield (2014) seems to have a similar idea when using the 'ought' of reasoning in a similar context. My presentation of this approach is closer to Schoenfield's.

Even without working out the specific details about the nature of special 'ought's, the approach provides an interesting way out of the detaching problem. If requirements employing special 'ought's are not requirements of morality or rationality, bad cases will not give rise to the kind of dilemma which characterizes the detaching problem. To see this, let us apply the Special Ought approach to Reflection:

> *Special Ought Reflection:* If you believe that you will rationally believe P in the future, you special-ought to believe P now.

An agent in a bad case irrationally believes her future self to believe P, and per Special-Ought Reflection, would be required to believe P now. However, if the special-ought is not the ought of rationality, the agent could be said to both be *rationally* required to abandon her irrational belief, and also be *special-required* to go along with it and believe P now. The agent would be in a situation where requirements from different domains conflict—which is nothing uncommon. For example, what prudence requires frequently conflicts with what rationality requires, and what rationality requires can conflict with what morality requires. Since the detaching problem relies on principles giving rise to conflicting requirements of the same normative domain, the Special Ought approach avoids it.

3.4 Hypothetical imperatives as world rankings

Some responders to the detaching problem have recently argued that hypothetical

imperatives should be understood as rankings of possible worlds.[10] On the *Ordering* approach, Commitment principles are claims about the ranking of certain worlds relative to a value. The following paraphrasing of Doxastic Commitment captures the point:

> *Ordering Doxastic Commitment:* Worlds in which you believe P in situation B and also believe Q rank rationally higher than worlds in which you believe P in situation B but do not believe Q.

The Ordering approach avoids the detaching problem by not requiring anyone to believe anything—thus not placing counterintuitive requirements on initially irrational agents.[11] All that such principles imply is that whether you believe P rationally or irrationally, worlds in which you also believe Q are ranked higher on the rationality scale than worlds in which you do not. So if you believe P rationally, you would occupy a rationally better world if you believed Q as well. And if you believe P irrationally, you would still occupy a rationally better world if you believed Q as well. Applied to Reflection, we get:

> *Ordering Reflection:* Worlds in which you believe that you will rationally believe P in the future and also believe P now rank rationally higher than worlds in which you believe that you will rationally believe P in the future

[10] There is a reading of this approach on which Commitment principle employ a special-ought—a second-best kind of ought. This would make the Ordering view a kind of Special-Ought view.

[11] See Silk (2014), Comesaña (2015).

but do not believe P now.

Ordering Reflection itself demands nothing of anyone. It provides no guidance with respect to what an agent should believe, whether or not the agent believes anything about her future self. It does, however, offer an ordering of worlds according to which some worlds are ranked higher (rationality-wise) than others. So without requiring anyone to do or believe anything, no detaching problem comes up.

3.5 What responses to the detaching problem share

The discussion in this section does not presume to provide a comprehensive mapping of all possible solutions to the detaching problem. Nevertheless, it is indicative of a feature that plausible solutions share. Defenses of principles facing a detaching problem try to preserve intuitions regarding both good and bad cases. Agents who start out rational really should form the further attitudes that their initial attitudes appear to commit them to. After all, that is the thought that the discussed principles set out to capture in the first place. Giving up on our intuitions in the good case would leave these principles hopelessly unmotivated, as no one should think that our attitudes commit us to further ones only when the initial attitudes are unjustified. On the other hand, agents who start out with unjustified attitudes cannot genuinely be required to simply go ahead and form further ones that their situations do not warrant, since this is how the detaching problem came up.

All four responses considered above are indeed compatible with these intuitions. The

Good Case approach keeps quiet about what agents in the bad case ought to believe, and only makes doxastic demands of agents in the good case. On the Wide Scope approach, Commitment principles impose certain prohibitions on everyone, but no demand to believe anything in particular on anyone. This leaves room for agents in the bad case to comply with wide scope principles by abandoning their initial irrational beliefs, and for agents in the good case to comply by forming the relevant further beliefs. On the Special Ought approach, Commitment principles put some kind of normative pressure on agents, but not the kind of pressure that gives rise to a dilemma that is internal to one normative domain. Lastly, the Ordering approach places no requirement or prohibition on anyone. All it does is offer a ranking of possible worlds relative to a value.

The shared feature I wish to draw attention to is this: all suggested approaches avoid the detaching problem by not rationally requiring agents in the bad case to adopt the further belief that their initial beliefs appear to commit them to. This means that if anyone is rationally required to adopt those further beliefs, it is agents in the good case. Combining this with the intuition that good-case agents are indeed rationally required to adopt those further beliefs that their initial ones commit them to, we get an important, albeit unsurprising, result:

> *Result:* Only agents holding rational beliefs are rationally required to adopt the further beliefs that their initial ones commit them to.

The approach that comes closest to explicitly stating Result is the Good Case approach. The other discussed approaches do not outright require anyone to believe anything, and are therefore not obviously committed to Result. However, any complete theory of rationality would be committed to Result, if such a theory were to respect our intuitions about the good and bad cases. So, irrespective of which (if any) of the answers to the detaching problem is correct, a complete story regarding what doxastic moves every agent must make would include Result. Answers to the detaching problem that stay silent about what agents should do, cannot reject Result without further argument.

Moreover, it is telling of Result's plausibility that none of the attempts to avoid the detaching problem disputes it. In principle, there could have been attempts to resolve the detaching problem that reject our intuitions about either the good or bad case. The absence of such attempts shows the authority of these intuitions, and bolsters Result.

Result is unproblematic when it comes to certain principles like Reflection. It seems right to say that only agents who have a rational belief that their future selves rationally believe P should believe P now. But Result is highly problematic when it comes to certain principles that have to do with higher-order evidence. The gist of the argument that reveals the problem is this: higher-order evidence appears to require us to revise a belief by suggesting that it is irrational. However, if only agents who hold the relevant belief rationally ought to change it in light of such information, then higher-order evidence will only trigger a belief revision requirement when it is importantly misleading—suggesting that the agent is irrational when she is not. In the next section I develop this argument and

explain why it shows that certain principles about higher-order evidence are dubious.

4 A problem for higher-order evidence principles

Consider two agents who share some body of evidence that supports P to degree d_1. Representing the good case, let agent S_g assess the evidence perfectly, and judge it to support P to degree d_1. Representing the bad case, let agent S_b assess the evidence poorly, and judge it to support P to degree d_2. Suppose both agents go on to assign credences d_1 and d_2 to P, according to their respective judgments. Then, a new bit of information is introduced. The agents are reliably informed of a bias that they share when it comes to assessing evidence about matters like P. They are told that some trustworthy researchers have observed their reasoning habits, and have concluded that they often overestimate the impact of their evidence on the likelihood of propositions like P by some degree.

We might think that when learning of their tendency to overestimate, agents should lower their degrees of belief by some amount. We might also think that a principle can capture the required rational response to learning of one's tendencies to overestimate, and suggest something like the following to that end:

> *Overestimation:* If you assign credence c to a proposition P, and you gain evidence of your tendencies to overestimate P, you should assign P a lower credence $c\text{-}x$.

Overestimation is a Commitment principle. The principle says that our starting attitude toward P commits us to a specific doxastic revision in the event that we gain evidence suggesting that we have overestimated P. The detaching problem surfaces when we consider what Overestimation would require of agents in a bad case, like S_b. Taken at face value, Overestimation implies that S_b is rationally required to move to a lower credence d_2-x. However, since S_b's attitude was unjustified to begin with, it cannot be that doing so would land S_b in a rational state.

Christensen's (2011) words about agents who start out irrational and then conciliate with their disagreeing peers clearly state the worry:

> [...] having taken correct account of one bit of evidence cannot be equivalent to having beliefs that are (even propositionally) rational, all things considered. If one starts out by botching things epistemically, and then takes correct account of one bit of evidence, it's unlikely that one will end up with fully rational beliefs. And it would surely be asking too much of a principle describing the correct response to peer disagreement to demand that it include a complete recipe for undoing every epistemic mistake one might be making in one's thinking. (Christensen 2011, p. 4)

Similarly, saying that d_2-x is a rational credence for S_b to have seems wrong. S_b initially misjudged her evidence as supporting P to degree d_2, and her irrational history is not expunged by her correct response to one piece of evidence about her overestimation tendencies. However, according to Overestimation, S_b is rationally required to move to

credence d_2-x. But if S_b goes ahead and does so, then that credence should be rational for her to hold—as it is incoherent to think that rationality requires we hold irrational credences. So Overestimation is open to a detaching problem.

To save Overestimation from the detaching problem, we could apply any of the four approaches discussed earlier:

> *Good Case Overestimation:* If you *rationally* assign credence *c* to a proposition P, and you gain evidence of your tendencies to overestimate P, you should assign P a lower credence *c-x*.

> *Wide Scope Overestimation: You ought to see to it that* when you assign credence *c* to a proposition P and you gain evidence your tendencies to overestimate P, then you assign P a lower credence *c-x*.

> *Special Ought Overestimation:* If you assign credence *c* to a proposition P, and you gain evidence your tendencies to overestimate P, you *special-ought to* assign P a lower credence *c-x*.

> *Ordering Overestimation:* Worlds in which you assign credence *c* to a proposition P, gain evidence your tendencies to overestimate P, and then assign P a lower credence *c-x*, *are rationally better than* worlds in which you

assign credence *c* to a proposition P, gain evidence of your tendencies to overestimate P, and do not assign *p* a lower credence *c-x*.

All four options indeed defend Overestimation from the detaching problem, as none requires agents in the bad case to lower their credence. But all of these principles face a separate problem.

At the end of the previous section I identified Result as an implication of any successful defense from the detaching problem. Result has it that agents in the bad case are not rationally required to form the further beliefs that their initial irrational beliefs appear to commit them to, whereas agents in the good case are. Applying this thought to Overestimation reveals a problem. Agents in the good case, by stipulation, have assessed their evidence rationally and have not overestimated. If only initially rational agents are rationally required to lower their credences when receiving evidence of their overestimation, then that evidence is guaranteed to be misleading whenever it prompts a credence-lowering requirement. Evidence of our having overestimated our evidence's support for P tells us that our evidence may in fact support some credence lower than the one that we assigned. As this would be incorrect of agents in the good case, such evidence will always mislead the only agents who ought to respond to it by lowering their credence. Any plausible version of Overestimation that both avoids a detaching problem and tells us how to respond to evidence of overestimation would require only agents who did not overestimate their evidence to lower their credences. Good Case Overestimation explicitly

requires this, and as I have argued, any properly detailed rule of how we should respond to evidence of our overestimation would. But we should not follow a rule that requires belief revision only in case one is being misled.

The misleading nature of principles like Good Case Overestimation is interesting. Being told that we overestimated P, when we in fact did not, may be non-misleading us with respect to P's truth-value. We may hold a perfectly rational high credence in P even though P is false—in which case any suggestion that we overestimated P would point in the direction of P's correct truth-value. However, information that we overestimated P would mislead initially rational agents in a different, important sense. It would mislead relative to whether the relevant belief is rational given the agents' other evidence—the very matter that appears to lend such evidence its prima facie evidential significance. So when we believe rationally, evidence that we believe irrationally would be misleading. The evidence would be anti-reliable relative to the proposition that it is explicitly about—i.e., relative to whether we believe irrationally—and so unreliable relative to whether P.

The problem is not unique to Overestimation, but affects many potential principles regarding the rational response to evidence about our rationality. The kind of information that we gain when we learn of our tendencies to be rational or irrational corresponds to what epistemologists often call higher-order evidence. Higher-order evidence is presented in the literature as evidence about the rationality of our beliefs, or evidence about what our evidence supports. For example, Richard Feldman (2009) understands higher-order evidence to be "Evidence about the existence, merits, or significance of a body of

evidence," and David Christensen (2010) describes higher-order evidence as "evidence that the evidential relations may not be as I've taken them to be." Following these conceptions, I take higher-order evidence to be evidence of a failure of rationality.[12] For instance, being told that you are drunk is evidence that the evidence you have may not support what you think it does. So, if we believe that we are okay to drive, such higher-order evidence would be indicative of our belief being, to use Christensen's terms, "rationally sub-par." Overestimation is one example of a principle trying to capture the rational response to a specific kind of higher-order evidence. Conciliationism, which I state in section (2), is another. The worry is that many potential principles regarding the required rational response to higher-order evidence would only require belief-revision of agents in the good case, for whom that higher-order evidence would be anti-reliable. More specifically, the problem comes up for principles that state the rational response to higher-order evidence, share the structure of Doxastic Commitment,[13] and are vulnerable to a detaching problem:

> *Higher-Order Commitment:* If you assign credence c_1 to a proposition P, and you gain higher-order evidence E, you should assign P a different credence c_2.

---

[12] This understanding leaves out evidence that confirms your rationality, and requires no obvious change in view. See Christensen (2010), p. 185-6. Nothing in the argument I give depends on this particular understanding.

[13] On Doxastic Commitment, when you believe P in some situation B, then you ought to believe Q.

The only agents, if any, who would be required to move from c_1 to c_2 would be ones for whom E is misleading. That is the general problem.

Principles that only require us to revise belief due to unreliable information are not reflectively endorsable. If we know that a principle like Overestimation only requires belief revision of agents who did not overestimate, our obligation to believe what our original evidence supports would require that we not let any information that we overestimated mislead us into believing differently than we do. We are rationally required to prevent unreliable information from affecting our beliefs as best we can—as long as it does not come at the expense of reliable information. And it is not just *known* unreliable information that we must treat this way. Information that can only affect what we should believe when it is unreliable deserves this treatment as well. Adhering to a principle like Overestimation would violate this requirement.

To summarize, the following is the problem with principles sharing the structure of Higher-Order Commitment. The only renditions of such principles immune to the detaching problem have the result that agents who believe rationally are the only ones who should change their beliefs due to higher-order evidence E suggestive of their irrationality. But if the only ones required to change their views are those who were not irrational, E would always be unreliable for them, and their commitment to believe what their original evidence supports would make it rational to dismiss the principle's command. In other words, the recommendations provided by principles sharing the structure of Higher-Order Commitment would be rationally dismissible, which suggests that these principles are

incorrect.

5 Responses that lead to dilemmas

One might worry that the arguments so far only show that principles about higher-order evidence cannot offer a guide for agents to follow, rather than show that those principles are implausible. The idea is that the fact that we cannot rationally follow the relevant kinds of principles, while believing in their correctness at the same time, does not obviously entail that the relevant principles are incorrect. In Kelly's words, "life is difficult",[14] so it could be an unfortunate part of epistemic life that we cannot follow the correct principles of rationality while being aware of them.

Principles that are not followable are suspect. For instance, an alleged rational requirement to believe what the ideal rational agent would believe may be beyond our ability to follow, and thus unsatisfactory. However, principles that are not rationally followable are even less plausible than those that we cannot follow as a matter of contingent fact. If an alleged rule of rationality is not rationally followable, then it means that our total evidence requires that we not follow it. In other words, rationality would require us to follow a rule that rationality requires us not to follow. That would not just make our epistemic life difficult, but it would make it impossible. If such a situation obtained then no matter what we were to do, we would violate some rational requirement. So, if alleged rules like Higher-Order Commitment obtained, they would be responsible for

[14] Kelly (2005), p. 180.

recurring rational dilemmas.

Dilemmas are just what we were trying to avoid all along. Recall the naïve Reflection principle discussed in the introduction. Reflection implies that if we irrationally believe that our future selves rationally believe P, we should believe P now. If that principle were true, we would be in a rational dilemma. Intuitively, we should fix our irrational beliefs, and stop thinking that our future selves rationally believes P, let alone believe P ourselves. The detaching problem is a worry about rational dilemmas arising all too frequently. If we were willing to grant that genuine rational dilemmas arise left and right, the detaching problem would be no problem at all, and the relevant principles would need no adjustments to begin with.

Lastly, there are two (related) ways to avoid detaching problems that I have not yet considered. On what we may call the *Total Evidence* view, true rational requirements specify how much weight we should assign to different parts of our total body of evidence. Rather than specifying the rational belief given some bit of higher-order evidence, the idea here is that each kind of evidence receives some weight, and the best we can do is find out how much weight evidence (like higher-order evidence) deserves. For instance, on such a view, evidence of our overestimation could get some percentage of the total evidential weight, and once we figure out what our other evidence supports we should employ the relevant modification given the weight that our higher-order evidence deserves. Kelly's (2010) view is one version of this approach, whereas more recently, Sliwa & Horowitz (2015) and Christensen (forthcoming) seem to be offering similar rules.

Total Evidence views fall victim to the worry discussed in the previous section. Rules that take our entire evidence into account would give higher-order evidence the exclusively destructive role of misleading us away from the beliefs that our other evidence supports. This, I argued, makes following such rules irrational. But there is a nearby view that may fair better. On what we may call the *Partial Evidence* view, the only true principles of rationality are those that tell us how to take correct account of particular bits of evidence. Some of Christensen's (2011) earlier quoted words about agents who start out irrational and then conciliate with their disagreeing peers hint at such an idea:

> [...] having taken correct account of one bit of evidence cannot be equivalent to having beliefs that are (even propositionally) rational, all things considered. (Christensen 2011, p. 4)

On the suggested view, some true principles of rationality only tell us the correct response to particular bits of evidence, and following those does not guarantee that the follower will end up with a rational belief.

The Partial Evidence view is not vulnerable to the main argument in this paper, and thus looks more viable. The view could have it that everyone must give higher-order evidence some evidential weight, even if doing so would not lead those who violate other rational rules to an all-things-considered rational belief. Since on this view following just one principle of rationality is not guaranteed to lead agents to an all-things-considered

rational belief, no detaching worries arise. The Partial Evidence view merely states a necessary condition on having an all-things-considered rational belief, according to which it is necessary to give higher-order evidence some evidential weight.

However, more would need to be said about what it means for it to be *necessary to give some evidential weight* to higher-order evidence, if it is to be a rule of rationality. Perhaps the following could work as a rough sketch of how such principles would go:

> *Higher-Order Partial Evidence:* It is rationally necessary to modify whatever credence S believes that P otherwise deserves by some degree depending on S's higher-order evidence.

Partial Evidence rules may provides guides to the necessary modifications of the credences that we form using our other evidence. Evidence of our tendencies to overestimate, for example, may trigger a requirement to modify our credence by lowering it, whereas evidence that our mental faculties are compromised could trigger a requirement to modify our credence by bringing closer to .5.[15]

But despite their apparent promise, such rules would also cause trouble. Take an agent who formed an irrational belief in P, and is told of her tendencies to overestimate P. Since she is irrational, her original evidence still requires her to change her view about P—

[15] Elga (ms.) gives as an example the case of Hypoxia—a condition that impairs reasoning without revealing to the reasoner that her reasoning is impaired.

perhaps from her irrational credence c_1 to the rationally required c_2. At the same time, a Higher-Order Partial Evidence rule would require that she use her irrational credence c_1 and modify from it, and not from c_2. At that point, the agent would be faced with a dilemma. Either she changes her initial judgment to c_2 and modify from there, or stick with her irrational c_1 and modify that. Both would claim to be necessary requirements of rationality, but they would conflict.

6 Conclusion

I have argued that attractive principles that capture the rational response to higher-order evidence will be hard to find. Plausible-sounding principles would have detachment problems to deal with, and the most promising ways of avoiding detachment problems lead to principles that require belief revision only of agents who were not irrational. This, I said, makes it rational to not follow these principles. In addition, I argued that the fact that it is irrational to follow an alleged rule of rationality is more than just a sad fact about the rule's inability to guide us, but rather makes the rule responsible for rational dilemmas. Lastly, I entertained two kinds of ways to avoid detaching problems, and argued that one is irrational to follow, and the other gives rise to inescapable rational dilemmas.

It is no happenstance that rules that require belief revision in light of higher-order evidence face the kinds of problems I have discussed. The arguments I raise reveal the fact that higher-order evidence either comports with what our other evidence already requires us to believe, or misleads. Once we notice this, it is not hard to generate objections to

alleged rules that claim to capture the correct response to higher-order evidence. Those rules will conflict with other rules of rationality if they have us stick with our past mistakes, or they will be hard to motivate if they constantly mislead us away from what our other evidence requires. This gives us reason to reconsider the intuitively significant evidential status of higher-order evidence.

Works cited

Broome, John. (1999). "Normative Requirements". *Ratio* 12, 398–419.

Christensen, David. (2007). "Epistemology and Disagreement: The Good News". *Philosophical Review* 116, 187–217.

–––– (2010). "Higher Order Evidence". *Philosophy and Phenomenological Research* 81(1), 185–215.

–––– (2011). "Disagreement, Question Begging, and Epistemic Self Criticism". *Philosophers Imprint* 11(6), 1–22.

–––– (2013). "Epistemic Modesty Defended". In Jennifer Lackey and David Christensen (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press. 77-97.

–––– (Forthcoming). "Disagreement, Drugs etc.: From Accuracy to Akrasia". *Episteme.*

Cohen, Stewart. (2013). "A Defense of the (almost) Equal Weight View". In Jennifer Lackey and David Christensen (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press. 98–120.

Comesaña, Juan. (2015). "Normative Requirements and Contrary-to-Duty Obligations". *The Journal of Philosophy* 112 (11): 600–626.

Elga, Adam. (2007). "Reflection and Disagreement". *Nous* 41(3), 478–502.

---- "Lucky to be Rational". Unpublished manuscript.

Enoch, David. (2010). "Not Just a Truthometer: Taking Oneself Seriously (But Not Too Seriously) in Cases of Peer Disagreement". *Mind* 119 (476): 953–997.

Feldman, Richard. (2007). "Reasonable Religious Disagreement". In L. Antony (Ed.), *Philosophers Without Gods*. Oxford University Press. 194-214.

---- (2009). "Evidentialism, Higher-order Evidence, and Disagreement". *Episteme* 6: 294–312.

Finlay, Stephen. (2010). "What *ought* probably Means, and why you can't detach it". *Synthese* 177: 67–89.

Kelly, Thomas. (2005). "The epistemic significance of disagreement". In John Hawthorne and Tamar Gendler (ed), *Oxford studies in epistemology*, volume 1. Oxford University Press, Oxford. 167-197.

---- (2010). "Peer Disagreement and Higher-Order Evidence:". In T. Warfield and R. Feldman (Ed.), *Disagreement*. Oxford University Press. 111–75.

---- (2014). "Believers as Thermometers". In Jonathan Matheson and Rico Vitz (Eds.), *The Ethics of Belief*, Oxford: Oxford University Press. 301–14.

Kolodny, Niko. (2005). "Why Be Rational". *Mind* 114 (455): 509–563.

Matheson, Jonathan. (2009). Conciliatory Views of Disagreement and Higher-Order Evidence. *Episteme* 6(3): 269–79.

Silk, Alex. (2014). "Why 'Ought' Detaches: Or, Why You Ought to Get With My Friends (If You Want to Be My Lover)". *Philosophers' Imprint* 14(7): 1-16.

Schoenfield, Miriam. (2015). "A Dilemma for Calibrationism". *Philosophy and Phenomenological Research* 91(2): 425-455.

Sliwa, Paulina. & Horowitz, Sophie. (2015). "Respecting All the Evidence". *Philosophical Studies* 172(11): 2835-2858.

Smities, Declan. (2012). "Moore's Paradox and the Accessibility of Justification". *Philosophy and Phenomenological Research* 85(2): 273–300.

Way, Jonathan. & Whiting, Daniel. (2015). "If You Justifiably Believe That You Ought to φ, You Ought to φ". *Philosophical Studies*, doi:10.1007/s11098-015-0582-2.

Way, Jonathan. (2010). "Defending the wide-scope approach to instrumental reason". *Philosophical Studies* 147: 213–233.

Weatherson, Brian. (ms.). "Do Judgments Screen Evidence?". Available at <http://brian.weatherson.org/JSE.pdf.>

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

Higher-Order Evidence Does Not Require Change in View

Abstract: Suppose we learn that we have a poor track record in forming beliefs rationally, or that a brilliant colleague thinks that we believe P irrationally. Does gaining such information require us to revise those beliefs whose rationality is questioned? When we gain information suggesting that our beliefs are irrational, one of two general cases obtains. In the first case we made no mistake and our beliefs are rational. In that case the information to the contrary is misleading. In the second case we indeed believe irrationally, and our original evidence *already* requires us to revise our belief. In that case, information to that effect is superfluous. Thus, information to the effect that our beliefs are irrational is either misleading or superfluous, and cannot justify belief revision.

1 Introduction

Some information we come across seems to bear on whether we believe things rationally. For example, we might learn that we tend to overestimate the likelihood of the weather being nice, or that we are hopelessly bad at forming rational beliefs about probabilities. We might also learn that others agree or disagree with our assessment of the evidence, or that they take our evidence to warrant a particular credence *c*.

The information we gain in these sorts of cases corresponds to what philosophers often call *higher-order evidence* (HOE). HOE is presented in the literature as evidence about the rationality of our beliefs, or evidence about what our evidence supports. For example, David Christensen (2010a) describes HOE as "evidence that the evidential relations may

not be as I've taken them to be" and Thomas Kelly (2010) understands HOE to be "evidence about the normative upshot of the evidence to which [one] has been exposed."

In what follows I argue that the information we receive in the kinds of examples above does not require revision of the doxastic attitudes whose rationality is in question.[1] I will use 'HOE' to refer to this type of information.[2] I will use 'original evidence' and 'other evidence' to refer to our evidential situation prior to some incoming HOE. I will refer to the doxastic attitude whose rationality is in question as the 'lower-level belief.' The discussion proceeds mainly in terms of coarse-grained belief, but is intended to apply equally well to a graded model of belief. Thus, 'belief' will be used interchangeably with 'doxastic attitude' and 'credence' unless otherwise specified. Given this terminology, the account I offer has it that no HOE requires revision of lower-level beliefs.

Throughout the paper I assume an evidentialist framework. On this framework, a belief is rational if and only if it accords with and is properly based on the agent's total evidence. I will also assume that rationality requires that we form the beliefs that our evidence supports, and in the right way, whether or not we are (or even can be) aware of what the

---

1 Some instances of HOE-acquisition involve more than just information about the rationality of our beliefs. For example, if an known epistemic superior, who has more evidence than we do, tells us that our evidence E misleadingly supports P, we learn both that they think *E supports P* and also that they think *P is false*. But by implying that the superior thinks P is false, such HOE carries with it some run-of-the-mill evidence against P, perhaps via some *Evidence of Evidence is Evidence* principle as Feldman (2014) suggests. I will focus the discussion on cases where the HOE does not bring with it such ordinary evidence.

2 Although I will not endorse either Christensen's or Kelly's conceptions of HOE, the arguments ahead go through equally well on their conceptions of HOE.

evidence supports. My goal is to show that there is a strong case to be made against the rationality of HOE-based belief revision[3] from a common evidentialist understanding of rational justification.[4]

To lay my cards on the table, here is a sketch of the argument to come. I argue that all instances of HOE fit neatly into two categories: *superfluous* and *misleading*. Superfluous HOE recommends that we should believe the very same thing our other evidence already requires us to believe. This kind of HOE changes nothing about what we should believe. Misleading HOE recommends something incorrect about what our original evidence supports. This kind of HOE lacks the main feature that makes HOE seem relevant to our lower-level beliefs, namely, an ability to correctly indicate what our evidence supports. No matter which of the two kinds the HOE is, we would know that either it does not change what we should believe (superfluous), or it is misleading about what our other evidence supports (misleading). Noticing this suffices for rationally refraining from HOE-based belief revision. If in the best scenario the HOE changes nothing about what we should believe, and in the worst scenario it misleads about what our other evidence supports, then HOE-based belief revision can only be motivated using prudential considerations. On the

---

[3] From here on, when I say HOE-based belief revision I will mean revision of the relevant lower-level belief.

[4] The list of phiosophers who have expressed sympathies to the thought that HOE requires lower-level belief revision is long. See, for instance, Feldman (2009), Kelly (2010), Christensen (2010a, forthcoming), Cohen (2013), Schechter (2013), Sliwa and Horowitz (2015), Schoenfield (forthcoming).

assumption that the requirements of (epistemic) rationality[5] are wholly independent of such considerations, we must put aside HOE when forming lower-level beliefs.[6]

In section 2 I offer a catalogue of all possible kinds of HOE, and show that they all fit into the two mentioned categories. In section 3 I argue that due to features of those categories, HOE-based belief revision is rationally forbidden. Section 4 is dedicated to explaining away the intuitive implausibility of such resistance to HOE, and section 5 to addressing worries that the account sanctions akratic doxastic attitudes. I conclude with an implication of the account for the prospects of certain total evidence rules in section 6.

2 Six Kinds of HOE

Any HOE can be characterized according to the following three parameters: *Valence*, *Correctness* and *Directionality*.[7]

Valence: HOE can be *positive* or *negative*. HOE is *positive* when it suggests that we believe rationally, and *negative* when it suggests that we do not believe rationally. For example, learning that a smart friend shares our belief given the same evidence is positive HOE,

---

5 When using the terms 'rational' and 'rationality' I mean to refer to their epistemic sense—the rationality of belief. I use 'epistemic requirements' and 'rational requirements' interchangeably. Later in the paper I discuss pragmatic rationality—the rationality of action—and make it explicit when I mean to refer to the pragmatic sense.

6 Even friends of pragmatic encroachment like Fantl and McGrath (2011) would grant that the credences we should form are independent of prudential considerations.

7 Although I have not found any, little depends on there being absolutely no exceptions to this claim. It is true of, at least, the vast majority of HOE. If the arguments to come only apply to the vast majority of HOE, that would be good enough.

whereas learning that she does not is negative HOE. Similarly, learning we have a perfect track record in forming rational beliefs is positive HOE, whereas learning we have a poor one is negative HOE. Simply put, HOE may suggest that we did or did not succeed in forming a rational belief given our evidence. So defined, we can gain negative HOE even in cases where we have no view about some matter. In those cases, any information suggesting that our evidence requires a particular credence *c* would be suggesting that we did not do a good job in forming a rational belief given our situation.[8]

Correctness: HOE can be *right* or *wrong*. What any HOE suggests about whether we believe rationally will either be true or not. When we believe rationally and gain positive HOE, or believe irrationally and gain negative HOE, the Valence parameter of the HOE is true. The same happens when we believe irrationally and gain negative HOE. In cases where the HOE's Valence parameter is true, the HOE is *right*. But the opposite can happen as well. We may gain information suggesting that we believe rationally when we in fact do not. We may also gain information suggesting that we believe irrationally when we in fact believe rationally. In those cases where our HOE's Valence parameter is false, the HOE is *wrong*. Thus, *rightly positive* HOE suggests that we believe rationally when we in fact do, whereas *wrongly positive* HOE suggests that we believe rationally when we in fact believe

[8] Of course, whenever we have no belief about some matter we would know that we did not form the rational belief given our evidence—assuming that rationality never requires that we abstain from having a view altogether. But on the way I characterize negative HOE, it is simply information that suggests that we did not form the rational belief. We can come across such information even when what it suggests is no news to us.

irrationally. Similarly, *rightly negative* HOE suggests that we believe irrationally when we in fact do, whereas *wrongly negative* HOE suggests that we believe irrationally when we in fact believe rationally. So defined, the HOE would be right even in cases where it points to a belief that is not supported by the evidence, as long as we indeed failed to form the rational belief. For example, the HOE is right when we believe a proposition that we should disbelieve, and the HOE suggests that we should suspend judgment. The Correctness parameter tracks the correctness of the HOE's Valence parameter.

Directionality: HOE can either be *directional* or *non-directional*.[9] If some HOE suggests that we have overestimated, underestimated, or accurately assessed our evidence, it is *directional*. This is because such HOE carries with it some indication of which belief is indeed required by our other evidence. Alternatively, if some HOE merely suggests that we believe irrationally and nothing more, it is *non-directional*. For example, learning that everyone thinks that we made some reasoning mistake is non-directional HOE, whereas learning that everyone thinks that our evidence supports a particular credence *c* is directional HOE.[10]

The three parameters specified apply equally well to a binary and to a graded framework for belief, and yield eight kinds of HOE:

---

[9] Christensen (2010b) introduces this distinction.

[10] Note that directionality fits both with a coarse-grained and a fine-grained account of belief. Learning we have overestimated our evidence is directional HOE whether we believe P or assign a high credence to P. Learning that we made some rational mistake or other is non-direction HOE whether we suspend judgment about P or assign medium credence to P.

HOE$_1$: Directional and wrongly negative.

HOE$_2$: Directional and rightly negative.

HOE$_3$: Non-directional and wrongly negative.

HOE$_4$: Non-directional and rightly negative.

HOE$_5$: Directional and wrongly positive.

HOE$_6$: Directional and rightly positive.

HOE$_7$: Non-directional and wrongly positive.

HOE$_8$: Non-directional and rightly positive.

However, notice that positive HOE will always be directional, since any indication that we believe rationally is an indication that our belief is rational *as is*. It follows that neither HOE$_7$ nor HOE$_8$ are possible kinds of HOE.

To show that none of the remaining six kinds of HOE requires belief revision, I first argue that they fit into two distinct categories, and then argue that the features of these categories justify resisting any HOE.[11]

[11] I do not deny that with suitably arranged background knowledge almost anything can be evidence for almost any proposition. On some quirky backgrounds, testimony that P is evidence against P, and a memory seeming as of Q is evidence against Q. But the existence of such backgrounds does not touch the interesting issues of the epistemic significance of testimony and memory. As I indicate in footnote 1, here I am interested in the epistemic significance of information about our rationality and what the our evidence justifies. So I put aside cases where incoming HOE concerns more than just our rationality or what our evidence justifies due to background knowledged.

2.1 HOE_1 and HOE_3 are misleading

HOE_1 and HOE_3 range over all instances in which the HOE is wrongly negative, i.e., wrongly suggesting that we have failed to believe rationally. Such HOE can be directional or non-directional, with varying degrees of specificity. For example, we may come across information suggesting that we made some mistake in reasoning, or that we underestimated the strength of our evidence, or that we tend to overestimate the strength of our evidence, or that we underestimated the strength of our evidence to some extent.

By definition, when we gain wrongly negative HOE, we hold the relevant belief rationally. So, HOE of this kind is guaranteed to be misleading in an important sense. It will be providing an incorrect indication of what our original evidence supports. This fact makes wrongly negative HOE misleading about the matter of what our original evidence supports—the very matter that makes HOE seem evidentially relevant to our lower-level beliefs in the first place.

2.2 HOE_4 is superfluous

HOE_2 and HOE_4 are both rightly negative, i.e., they suggest that we believe irrationally when we in fact do. Such HOE can also be directional or non-directional, and to varying degrees of specificity. Examples of what such HOE looks like are the same ones given in the previous sub-section for wrongly negative HOE.

When rightly negative HOE is non-directional (HOE_4), it correctly indicates that we do not believe rationally, but stays silent on what the rational belief is. Consequently, any

move to a new belief based on such HOE would be arbitrary. This remains true even if, on the basis of such evidence, we suspend judgment, which may seem like a non-arbitrary response to that kind of HOE. Suspension of judgment, like any other doxastic attitude, is permitted only when the balance of evidence goes a particular way. Not knowing how to properly weigh the evidence for and against a proposition no more justifies suspension of judgment than it justifies moving to the exact opposite belief to one's own. Suspension of judgment is not a doxastic safe-zone that we may occupy while unsure about what our evidence supports, and where we are shielded from rational criticism. If we want to resort to a doxastic safe-zone, perhaps we could stop having a view on the relevant matter.[12] But this too cannot be what non-directional HOE requires when it tells us of our failure to form a belief rationally. The issue of whether we should have a view about a matter does not fall within the jurisdiction of rational requirements.[13] Rational requirements instruct us that *if* we form a view on a matter, it be the appropriate one. So non-directional and rightly negative HOE does not change what we should believe.

Importantly, there is no sense in which we would be *more required* to form the rational belief after gaining information to the effect that we believe irrationally. Requirements are binary, either applying to an agent in a given situation or not. One cannot be more

---

[12] There is a worry here that such a move may not be psychologically possible.

[13] Cases where we know we are guaranteed to believe irrationally no matter what may look like ones in which rationality requires we have no belief. But on an evidentialist framework, rationality and irrationality are properties of beliefs, and cannot be attributed to absence of belief. I therefore take such cases to show, at most, that sometimes we are doomed to fail to believe rationally upon trying.

required or less required to believe P. Since our original evidence already requires us to form a rational belief, we still should (and no more than we previously did) if we learn that we currently believe irrationally.

What I say here does not imply that an agent who already believes irrationally and also ignores HOE about her having made a rational error is *as irrational* as an agent who holds an irrational belief but never receives such HOE. It may be that irrationality comes in degrees, so that when we ignore information about having made a mistake we are more irrational than when we merely make the mistake. Perhaps this natural thought is due to the fact that other requirements would be violated in those apparently worse cases—maybe some requirement to never ignore information. However, we should resist the move from the thought that irrationality comes in degrees to the thought that the rational requirements themselves come in degrees. The separation seems natural in other domains. For instance, the law (and morality) requires us not to drive while heavily intoxicated whether or not a friend points out that we have been drinking excessively. Receiving more evidence that we are violating a requirement need not imply an increase in the normative force of the relevant requirement, even though violating the requirement in such situations may be worse—e.g., in virtue of other requirements violated. So even if failure to meet a requirement can be more or less severe, it does not follow that one can be more or less required to meet it.

From this we should conclude that non-directional and rightly negative HOE (HOE_4) does not change what an agent is required to believe. We should correct our doxastic

delinquencies before learning about them just as much as we should after learning about them. What we should believe before and after such HOE is the same, making HOE_4 superfluous.

2.3 HOE_2 is superfluous or misleading

Unlike HOE_4, HOE_2 does not only suggest (rightly) that we believe irrationally, but also tells us something about the direction in which we must revise. In so doing, this kind of HOE can get certain things wrong. There are cases in which the HOE suggests that we have overestimated our evidence when we in fact underestimated it (and vice versa). There are also cases where the HOE correctly suggests that we have overestimated or underestimated our evidence, but incorrectly indicates the degree to which we have. So even when right in suggesting that we believe irrationally, our HOE may still get more specific details wrong. Directional and rightly negative HOE can be *accurate* (HOE_{2a}) or *inaccurate* (HOE_{2i}), depending on whether it correctly or incorrectly indicates the direction (or extent) of the mistake made.

Accurate, directional and rightly negative HOE (HOE_{2a}) accurately points in the direction of (or precisely to) the rationally required belief, when we really did fail to reach it. Since it is accurate, it recommends the exact same belief that our original evidence does, and is thus superfluous.

Inaccurate, directional and rightly negative HOE (HOE_{2i}) is misleading. By definition, inaccurate feedback about which mistake we made is incorrect with respect to the mistake

made. This information thus provides a misleading indication of what attitude is rational on our other evidence. So when inaccurate, HOE_2 is misleading, and when accurate, it is superfluous.

2.4 HOE_5 and HOE_6 are superfluous or misleading

HOE_5 and HOE_6 capture all instances in which the HOE is positive, i.e., suggesting that we hold our belief rationally. Such information intuitively warrants no belief revision. It would be odd to think that positive feedback about our belief's rationality might require that we abandon the belief. This is already some reason to think that neither HOE_5 nor HOE_6 require belief revision.

Notice that HOE_6 suggests that we hold our belief rationally when we in fact do, making it superfluous. If such HOE supports any belief, it must be the very same one that our other evidence does, and that we already hold. Meanwhile, HOE_5 suggests that we hold our belief rationally when we in fact do not. HOE_5 will therefore be misleading, recommending a belief different than the one supported by our other evidence. It follows that positive HOE is either superfluous or misleading about what our other evidence supports.

3 The Moral For Higher-Order Evidence

We can now see that the six kinds of HOE fit into two general categories. HOE_{2a}, HOE_4 and HOE_6 change nothing with respect to what our original evidence already requires that

we believe, and are thus superfluous. HOE_1, HOE_{2i}, HOE_3 and HOE_5 provide an incorrect indication of what our original evidence supports, and are thus misleading. While we would often not know which of these two categories our HOE falls under, we would know that it either does not change what we ought to believe, or misleads about what our original evidence supports. But if we know that our HOE can only change what we should believe when it is misleading about what we should believe, it is hard to see why we should ever let it affect our beliefs.

Misleading HOE has much in common with other bad HOE. Disagreement from known epistemic inferiors may provide information about what our shared evidence supports, but is not a reason to change our view on the disagreed upon issue. Epistemic inferiors are known to be no good at telling what our other evidence supports regarding a proposition P, and by extension no good at telling whether P. That is why when we know we face one, we should not take the offered input as reason to change our view of P. Of course, unless we know that we face some misleading or otherwise bad input, it may be irrational to form our beliefs about P without it. Misleading information that we do not know is misleading often requires us to revise our beliefs. For this reason, when we do not know whether our HOE is misleading, it is tempting to think we should revise our beliefs at least to some extent in response—in case it is not misleading. But here is where things are different with HOE. If we know that the only alternative to its being bad at telling what lower-level belief we should have is that it changes nothing about what lower-level belief we should have, we should form our views about the relevant matter using our other evidence

alone.

That we should dismiss disagreement that comes from a known epistemic inferior seems clear. But what should we do with advice from a *potential* inferior? Imagine that we rationally believe Smith to be our epistemic equal, and that Smith offers us some advice regarding what our evidence about P supports. Suppose we then learn something about the situation: either the offered advice changes nothing about what we should believe about P, or Smith is in fact vastly intellectually inferior. In this case we do not know that Smith is an epistemic inferior, but we may as well treat him like one. His advice can only change what we should believe in the event that it is bad advice, leaving us with no reason change our beliefs due to his input (and enough reason not to). The characteristics of the case, however, are shared by all cases of HOE. HOE can only change what we should believe when it is bad advice about what we should believe. Knowing this is sufficient for rationally not taking it.

To further motivate the thought, consider a moral analogue:

> **Directions**
>
> Jones morally ought to (do her best to) reach Rome. She is given a set of accurate directions D_1, which she knows is both accurate and easily within her ability to follow. Before trying to follow D_1, she is provided with an alternative set of directions D_2, which she knows to be one of two kinds: either D_2 contains accurate and easier-to-follow directions that lead to

Rome, or it contains inaccurate directions that lead to somewhere else.

Assuming Jones has only one attempt to make it to Rome and can use only one set of directions, may she opt to use D_2? It seems clear she must not. If reaching Rome (or doing her best to reach Rome) is really all that is required of Jones, and if D_1 accurately directs her there, she must follow it alone. Following D_2 would amount to sacrificing her possessed and perfectly accurate directions D_1, which direct her to doing as she ought to, in exchange for convenient directions at best, and misleading directions at worst. Doing so would not count as Jones's best attempt to fulfill the obligation, and so following D_2 is morally impermissible.

In both Directions and in cases of HOE, the agent has perfectly accurate directions for doing what she ought to do. In Directions, it is D_1. In cases where we are presented with HOE, it is our original evidence—since assessing it correctly is guaranteed to lead us to the belief it supports. In both Directions and cases of HOE, an opportunity presents itself to forgo the perfectly accurate set of directions the agent possesses, in exchange for another, at best more convenient and at worst worthless set. In both, making the exchange is normatively forbidden. If Jones were to use D_2 rather than her reliable D_1, she would be morally criticizable. If we rely on HOE to revise our lower-level beliefs instead of on our other evidence alone, we would be rationally criticizable. This verdict seems to hold even in the event that D_2 (or our HOE) is more likely to be of the accurate and easy to follow kind than of the misleading kind.

There is an obvious disanalogy between Directions and cases of HOE. Directions is set up so that the agent is easily able to do as she ought by following D_1, whereas we often cannot figure out what our evidence supports (and thus what we ought to believe). This difference may lead some to think that making use of HOE is rationally permissible, because it helps us obey the rational requirements imposed by our original evidence. In a sense, this is true. Prudentially speaking, in order to fulfill the goal of forming the belief that our original evidence justifies, we would do well to follow our HOE where it leads, despite risk of it misleading us. After all, our HOE is frequently a better guide to what our original evidence supports than we are. But the prudential benefits of using HOE can compromise our intuitions, and make it seems like rationality requires that we use it in doxastic revision. Facts about what steps we can take to improve our odds of reaching the belief required by our evidence do not figure into rational requirements. For example, getting a good night's sleep would improve these odds, but is not rationally required. It is easy to conflate the requirement to believe what the evidence justifies with the requirement to do one's best to believe what the evidence justifies. This is because doing one's best to believe what the evidence supports usually helps with believing rationally. But these requirements come apart. This key observation is necessary for showing that HOE-based belief revision can only be a kind of prudential bet hedging, rather than anything that rationality requires.

To sum up, our original evidence acts as a perfectly accurate guide to what we ought to believe. If we form our lower-level beliefs by incorporating HOE, we would be sacrificing

this perfectly accurate guide in exchange for a potentially inaccurate one. This observation highlights the fact that the only thing HOE-based belief revision has going for it is its ability to help us believe what we already should. Since rationality does not require we do out best to believe rationally, it does not require HOE-based belief revision.

In the next section I say more about why resisting our HOE feels like an irrational move even though it is not. After that, I address the worry that the view sanctions *epistemic akrasia*—i.e., maintaining a belief while believing that it is irrational.

4 Residual Incredulity: Prudential vs. Rational Belief

The implications of the argument presented are surprising. It is hard to believe that we should really take no amount of higher-order evidence to require belief revision. When all of the experts tell us that we have overestimated the strength of our evidence, or when we learn that we have been dosed with a powerful reason-distorting drug, it seems like the only way to take that input seriously is by significantly revising our beliefs.

In the previous section I raised the possibility that this seeming stems from our confusing prudential belief revision with respect to the goal of acquiring rational beliefs, with rational belief revision. The explanation fits nicely with, and is enhanced by, a distinction that Joshua Schechter (2013) draws in the context of HOE. Schechter distinguishes between the requirements of *epistemic justification* and the requirements of *epistemic responsibility*. We can draw a similar distinction between rational belief and responsible belief. Accordingly, we may evaluate HOE by whether it changes the rational

requirements that we are under, and by whether it changes the responsibility requirements that we are under. On the account I offer, HOE does not affect the former, for it can add nothing (non-misleading) to what we should already believe given our other evidence. But perhaps, in order to count as responsible belief holders, we may not put aside information about our having made some mistake. Perhaps responsibly responding to this kind of information involves checking our reasoning, looking for errors,[14] or simply following our HOE in order to maximize our odds of reaching the belief our original evidence supports. Failure to respond to HOE in these ways may entail a violation of a responsibility requirement, even though it does not entail a violation of an rational requirement. On my view, then, rational belief need not be responsible (or prudent) belief.[15]

The intuition that resisting HOE is irrational likely originates in taking the requirements of responsible/prudent belief formation to be rational requirements. If the former were kinds of epistemic requirements, compliance with them would be necessary for believing rationally. Then, in cases of HOE, an agent would not be rational in maintaining her view until she respected the HOE by sufficiently rechecking her view, or doing whatever else might count as the responsible response. But we need not grant that requirements of responsible belief formation are so related to rational requirements. If we deny that they are, we can explain why resisting HOE does not interfere with believing

---

[14] See Schechter (2013).

[15] Schechter (2013) takes irresponsibility to undermine rationality, and would thus reject the position defended here.

rationally.[16]

Requirements of responsible belief formation still have some significance relative to rational belief. Good habits of belief formation promote rational belief acquisition. Checking our reasoning for mistakes helps us avoid irrationality, and since avoiding irrationality *is* rationally required of us, it might seem that checking our reasoning is too. Recall, though, that many other things facilitate avoidance from forming irrational beliefs. Getting a good night's sleep is one example I have mentioned. It is nothing more than a contingent fact about us that we often fall short of fulfilling rational requirements, and various courses of action (some, doxastic) can minimize the occurrence of this phenomenon. Yet it is not rationally required to abide by the requirements of responsible belief formation, just as is not rationally required to get plenty of sleep. It is merely useful to do so. The rationality of a belief depends on the belief's accordance with our evidence. The rationality of a belief does not depend on our taking actions that improve our odds of believing in accordance with our evidence.

We should therefore resist the intuitive strangeness of the thought that our beliefs can be perfectly rational despite, for example, many experts falsely telling us that we believe irrationally. It may well be that if this were to happen we would all, as a matter of fact, give

---

[16] Schechter (2013) points out that here lurks an akrasia-related concern. If we know that HOE imposes a responsibility requirement to double check our reasoning, then even when we hold P rationally, negative HOE should make us think we are holding it irresponsibly. But, intuitively, if we rationally believe that responsibility requires we double check our reasoning, it is because we believe we may have made some reasoning error along the way, and that we may be holding P irrationally. I address such concerns in the next section.

up our beliefs. It would indeed be a good move, practically speaking, to change our view when the experts tell us it is irrational—if our goal is to have beliefs that accord with our original evidence. And it may even be that we should, in some sense, *excuse* someone who does follow their HOE.[17] But as I argued, we are not rationally required to take the steps that maximize the odds of having such beliefs. Moreover, those more extreme examples where the intuition is strongest are quite rare—the experts would have to be mistaken or conspiring to fool us. This makes for a further reason to think that our gut reaction to such cases is unreliable.

5 Epistemic Akrasia

I have argued that HOE does not require revision of lower-level beliefs, that is, those beliefs the rationality of which our HOE concerns. One important implication of this view is that we should sometimes believe P despite possessing lots of HOE that suggests we should not. Call this implication Steadfast:

> *Steadfast:* If we believe P rationally and then receive HOE suggesting our belief is irrational, we are (still) required to believe P.

I have kept quiet about how HOE affects the beliefs that we should have regarding whether our beliefs are rational—i.e., our *higher-level* beliefs. Indeed, HOE often seems to

[17] See Littlejohn (forthcoming) on epistemic excuses.

require revising higher-level beliefs. For instance, an expert telling us that we believe irrationally seems to provide excellent reason to think that we believe irrationally. Just as testimony that P is normally evidence for P, testimony that we believe irrationally looks like good evidence that we believe irrationally. So, strong enough negative HOE should require believing that the corresponding belief of ours is irrational. Call this natural thought Higher-Level:

> *Higher-Level:* Strong HOE suggesting a belief is irrational requires believing that the belief is irrational.

Steadfast and Higher-Level combined are in tension with the enkratic constraint:

> *Enkrasia:* Rationality forbids having a doxastic attitude *d* toward P while also believing that having that attitude is irrational.

According to Enkrasia, agents are never rationally required both to believe a proposition and to believe that believing it is irrational. But if we are sometimes required to retain a belief that P despite HOE that it is irrational (per Steadfast), and also required to believe that believing P is irrational (per Higher-Level), then we are sometimes required to believe P and also believe that believing P is irrational (contra Enkrasia). So Steadfast, Higher-Level

and Enkrasia are jointly inconsistent, and we must give at least one up.[18] Since Higher-Level would seem to follow from any reasonable account of testimony, Enkrasia and Steadfast appear to be in trouble.

Rejecting Enkrasia comes at a high cost. Beyond its intuitive appeal, convincing arguments against its rejection are readily available. Horowitz (2014) argues that views that allow akratic combinations of attitudes "license patently bad reasoning and irrational action".[19] For example, consider the following version of a case given by Horowitz:

> Detective correctly assesses her evidence E and comes to believe that Smith committed the crime. She then acquires strong but misleading HOE suggesting that E supports that Jones is the perpetrator. She reasons as follows: *Smith is guilty, as I believe based on E alone. But my HOE suggests that E does not support Smith's guilt. So the evidence E on which I'm basing my belief must be misleading. Nevertheless, based on E, I believe that Smith committed the crime and should have him arrested!*

Two aspects of this case are disconcerting: First, Detective's inference that her evidence E is misleading, and second, Detective's maintaining her view despite basing it only on E—evidence she considers to be misleading—and despite believing that this belief is irrational.

---

[18] Others have discussed versions of this inconsistency. See Christensen (2013), Horowitz (2014), and Worsnip (2018).

[19] Horowitz (2014), p. 11.

Any view that countenances akratic attitudes, and allows (as we plausibly should) that agents may reason and act based on their rationally held beliefs, is committed to these problematic implications.

In light of the initial plausibility of Enkrasia and Higher-Level, and given their inconsistency with Steadfast, the proponent of Steadfast seems to owe a story about which of these two principles we should reject and why. A few ways of rejecting Enkrasia and Higher-Level are found in the literature, and may be combined with the account defended so far.[20] I now argue that Enkrasia and Higher-Level are already in tension with one another. This argument is made possible given the following level-connection principle:

> *Self-Intimation:* When we have (or lack) justification to believe P, we have justification to believe that we have (or lack) justification to believe P.[21]

I now argue that Self-Intimation requires us to rationally refrain from revising higher-level beliefs due to HOE, thereby undermining Higher-Level. I then show how accepting Enkrasia commits one to Self-Intimation.

Let S_n stand for an agent's situation, and let Jp stand for the proposition that *we have justification to believe P*. It follows from Self-Intimation that for any P and S_n, we either have

---

[20] For instance, for a rejection of Higher-Level see Titelbaum (2015), and for a rejection of Enkrasia see Weatherson (ms.), Williamson (2011, 2014), Coates (2012), Wedgwood (2012), and Lasonen-Aarnio (2014).

[21] I borrow this from Smithies (2012).

justification to believe it and to believe that we have justification to believe it, or we lack justification to believe it and have justification to believe that we lack it. We are thus necessarily in one of two possible kinds of situations: either both Jp and JJp obtain (situation S_1), or both ~Jp and J~Jp obtain (situation S_2).

If HOE is to require a change in our higher-level beliefs, it must weigh in on whether we have justification to believe P in our situation S_n. In other words, HOE must support either Jp or ~Jp. So HOE can either suggest that Jp obtains in S_n or that ~Jp obtains in S_n. Let HOE+ stand for the former and HOE- stand for the latter. From these stipulations, four possible HOE/S_n combinations emerge. We could gain HOE- in S_1, HOE+ in S_1, HOE- in S_2, or HOE+ in S_2. These four possible combinations correspond to cells A through D in the table below:

	HOE- (~Jp in S_n)	*HOE+* (Jp in S_n)
S_1 (Jp & JJP)	A	B
S_2 (~Jp & J~JP)	C	D

Now consider the higher-level proposition that *we have justification to believe P* (Jp). In scenario A our HOE suggests that we lack justification for P in our situation. Since the situation there is S_1, this is false, and the HOE is misleading relative to Jp. In scenario B our HOE correctly suggests that we have justification for P in our situation. Since the situation there is S_1, we already have justification to believe we have justification for P in the form of JJp. Thus, the HOE in scenario B is superfluous for justifiably believing Jp. In

scenario C our HOE correctly suggests that we lack justification for P in our situation. Since the situation there is S_2, we already have justification to believe we lack justification for P in the form of J∼Jp. Thus, the HOE is again superfluous in this scenario. In scenario D our HOE suggests that we have justification for P in our situation. Since the situation there is S_2, this is false, and the HOE is again misleading. So, relative to Jp, the HOE is misleading in scenarios A and D, and superfluous in scenarios B and C.

Earlier in this paper I argued for a similar categorization of HOE. There, the proposition of interest was P. I argued that HOE is either superfluous for having the rational belief about P, or misleading relative to Jp (which makes HOE a bad guide to whether P). From that I concluded that we should not let HOE affect our beliefs about P, and our lower-level beliefs in general. The argument here is that, given Self-Intimation, HOE is either superfluous for having a rational belief about Jp, or misleading relative to Jp. Here, and for the same reasons as before, the upshot is that we should not let HOE affect our beliefs about Jp and our higher-level beliefs in general. Knowing that our HOE is misleading or superfluous relative to Jp gives us reason to form our belief about it using the other justification we already have. The HOE can only change what we should believe about Jp by misleading us, and the potential convenience it offers is a prudential consideration at best.

In one respect, the argument against HOE-based higher-level belief revision proceeds more smoothly than the one against HOE-based lower-level belief revision. The misleading/superfluous accusation here is indexed to a single proposition Jp. But there is

also a respect in which the argument here proceeds less smoothly. If we think justification for higher-level beliefs comes in degrees, what I count as superfluous HOE can simply be considered *additional* evidence in favor of the correct higher-level proposition. This could not be said of superfluous HOE relative to a lower-level belief that P. Whatever doxastic attitude our original evidence requires with respect to P, HOE cannot make us more required to have that attitude. But what I categorize as superfluous HOE relative to Jp is HOE that bolsters the justification we already have to believe Jp (or ~Jp). So it is fair to wonder why the possibility that our HOE is additional evidence in these cases does not give HOE some say about the rational higher-level belief.

We may say two things here. First, when HOE adds to the justification that we already have for the correct higher-level belief, it does not change what our *coarse-grained* higher-level belief should be. If we already had justification to believe Jp (or ~Jp), any HOE that adds to that justification does not change what our coarse-grained belief about Jp should be. This seems like a good reason to form our coarse-grained higher-level beliefs based on our other justification alone. HOE can only change what coarse-grained beliefs we should have by misleading us about whether Jp.

Second, principles specifying what combinations of attitudes are akratic (like Enkrasia) often forbid combinations of any lower-level doxastic attitude (coarse or fine-grained) with a conflicting *coarse-grained* higher-level belief that the lower-level attitude is irrational.[22] This

[22] On Horowitz's (2014) "an *epistemically* akratic agent believes something that she believes is unsupported by her evidence." On Titelbaum's (2015) akratic principle "no situation rationally permits any overall

is not happenstance. Having an attitude while believing that it is irrational strikes many as patently incoherent, whereas having an attitude while being uncertain about whether it is rational is not clearly so.[23] Intuitions are murky regarding precisely which combinations of lower-level and higher-level doxastic attitudes should count as akratic when higher-level beliefs are taken as fine-grained. Surely a lower-level credence of .4 and a maximal higher-level credence 1 that .4 is rational counts as a non-akratic combination. But intuitions about other combinations vary. A situation where we assign .7 to the proposition that .4 is the rational credence to have in P does not seem *that* akratic, although more akratic than one where we assign .9 to .4 being rational. By contrast, a situation where we assign .3 to the proposition that .4 is the rational credence to have toward P does seem akratic. So treating higher-level beliefs as fine-grained makes stating the correct akratic principle a complicated matter. Yet without a statement of such a principle, it would be hard to pose the main concern of this section. So insisting on a graded framework for higher-level beliefs interferes with opposing Steadfast on grounds of violating an enkratic constraint.

The argument against HOE-based higher-level belief revision rests on the plausibility of Self-Intimation, given the assumption that Enkrasia is true.[24] Declan Smithies (2012, 2016)

---

state containing both an attitude A and the belief that A is rationally forbidden in one's current situation."

[23] Huemer (2011) argues that such a combination of attitutes is irrationa. Hazlett (2012) argues that we may sometimes rationality hold beliefs while suspending judgment about whether they are reasonable.

[24] Kelly (2010) offers some Enkrasia-independent reason to think that Self-Intimation and similar level-connection principles obtain. He argues that recognition of the import of our evidence is often the reason we form the paradigmatically rational beliefs that we do. It is not a major leap from this

provides an argument that establishes a strong link between the two. Smithies argues that any view interested in avoiding the epistemic version of Moore's paradox must accept Self-Intimation. The Moorean propositions that Smithies targets as to-be-avoided include *p and I do not have justification to believe p*, and similar propositions that proponents of Enkrasia disapprove of. Enkrasia has it that rationality forbids having any doxastic attitude *d* toward P while also believing that having *d* toward P irrational. On Smithies's view, if our situation justifies a doxastic attitude *d* toward P but does not justify the belief that it justifies *d*, then it must either justify suspension of judgment or disbelief that it justifies *d*.[25] Yet, whether we should disbelieve that our situation justifies *d* or suspend judgment on the matter, it would violate Enkrasia to maintain *d* at the same time. Both suspension of judgment and disbelief that we have justification for *d* are ways of believing that we lack justification for *d*, and thereby that having the attitude is irrational. So it must be that whatever attitude toward P our situation justifies, it also justifies the corresponding higher-level belief that it justified *d*, rather than disbelief or suspension of judgment about the matter. Accepting Enkrasia, therefore, commits us to Self-Intimation.

Self-Intimation is what enables the argument against HOE-based higher-level belief

---

observation to thinking that a necessary requirement on rational belief is that agents recognize what their evidence supports. Moreover, as Kelly argues, recognition that such-and-such entails justification to believe that such-and-such. So when we ought to believe P and do so rationally by recognizing that our evidence supports it, we may be justified in believing that our evidence supports it.

[25] Like Smithies, I assume here that our total evidential situation "is exhaustive in the sense that, for any proposition, one has justification either to believe, to disbelieve, or to withhold belief in that proposition."

revision to proceed. As a result, Higher-Level, which states we must change higher-level beliefs due to HOE, is undermined. Importantly, Self-Intimation can be motivated without presupposing Steadfast. Enkrasia and Steadfast are enough to undermine Higher-Level, and so presupposing Steadfast would beg the question. But if Enkrasia alone commits us to Self-Intimation, we can undermine Higher-Lever without appeal to Steadfast. The objection to Steadfast saying that it forces us to reject one of two intuitively plausible principles is thus diffused. We must reject either Enkrasia or Higher-Level no matter what view we hold about the significance of HOE to our lower-level beliefs.

6 Total evidence principles

The observation that HOE is always superfluous or misleading gives rise to related arguments we can employ at other points in the debate. Specifically, some non-steadfast rules that claim to capture how we should let HOE affect our lower-level beliefs seem especially dubious once we realize the nature of HOE. Consider, for instance, the following schema for such a rule:

> *Total Evidence:* The credence in P that S should have is the one that S's original evidence supports, modified in some way by what her HOE supports.

Let a total evidence rule be any rule that offers a way to modify what our original evidence

supports according to our HOE.[26] In doing so, total evidence rules give HOE the questionable role of either changing nothing about, or misleading us away from, what our original evidence requires us to believe. This is suspect.

The complaint is not that total evidence rules require that we form a belief based on our original evidence and then allow HOE considerations to mislead us away from it. I am not too worried about the followability prospects of such rules (although that is a fair thing to worry about). Rather, the complaint is that according to such rules, HOE only affects what we should believe when it is misleading. This fact opens these rules up to competition from strictly simpler and less demanding rules, which do not mislead. Relative to Total Evidence, for instance, such a competing rule can be reached simply by striking out the inclusion of any sensitivity to higher-order evidence:

> *Less Demanding, Non-misleading Total Evidence:* The credence in P that S should have is the one that S's original evidence supports. ~~modified in some way by what her HOE supports.~~

Any agent who can follow Total Evidence can follow Less Demanding, Non-misleading Total Evidence. Moreover, the latter leads us exactly where we wanted to be all along, namely, at a belief that accords with our original evidence. What lends HOE its apparent

[26] Christensen's (forthcoming) Idealized Thermometer Model and Sliwa & Horowitz's (2015) Evidential Calibration can arguably be understood as such rules.

evidential significance is its claimed ability to direct us to what our original evidence supports. But if HOE can only affect what we should believe by misleading us away from what our original evidence supports, it loses the only evidential appeal it ever had. So by leaving out HOE from consideration altogether, we end up with a simpler, easier to follow, non-misleading rule, and thus one that is much more attractive than total evidence rules that give HOE a role in affecting lower-level beliefs. Since there is no reason to think the correct rules of rationality would be trickier than they need to be, the less demanding and non-misleading rule is a more plausible rule of rationality than Total Evidence. But of course, the less demanding and non-misleading rule is empty.

7 Conclusion

The steadfast position defended in this paper is, by nature, a radical one. In support of this position I have argued that higher-order evidence at best changes nothing about what our other evidence already requires, and is at worst misleading about what our other evidence requires. So whatever our original body of evidence is, *that* body of evidence determines what is rational for us to believe about different propositions.

The account has an interesting consequence regarding the nature of HOE. HOE turns out not to be evidence for or against the lower-level belief it concerns. HOE still seems relevant in its triggering of a responsibility requirement to check our assessment of the evidence—as responsible inquiry involves checking for errors when suspicion of error becomes salient. But whenever we make a rational error, our evidence already requires

belief revision. All that HOE could do in such a case is call attention to what we are already required to believe.

Perhaps even more surprising is that for those who are sympathetic to the akratic principle, HOE turns out not to be evidence for or against the higher-level belief it *directly* speaks to. But if HOE is neither evidence about lower-level matters nor about higher-level matters, then it is hard to see in what respect it has any evidential value.

Works cited

Christensen, David. (2010a). "Higher-Order Evidence". *Philosophy and Phenomenological Research* 81(1): 185–215.

–––– (2010b). "Rational Reflection". *Philosophical Perspectives* 24: 121-140.

–––– (2013). "Epistemic Modesty Defended". Jennifer Lackey and David Christensen (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 77–97.

–––– (Forthcoming). "Disagreement, Drugs, etc.: From Accuracy to Akrasia". *Episteme*.

Cohen, Stewart. (2013). "A Defense of the (almost) Equal Weight View". Jennifer Lackey and David Christensen (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press. 98–120.

Comesaña, Juan. (ms.). "On An Argument Against Immediate Justification". Available at <http://comesana.web.arizona.edu/files/on-an-argument-against-immediate-

justification.pdf>

Feldman, Richard. (2014). "Evidence of Evidence is Evidence". In *The Ethics of Belief*, ed. John Mattheson and Rico Vitz. 284–99.

–––– (2009). "Evidentialism, Higher-order Evidence, and Disagreement". *Episteme* 6: 294–312.

Hazlett, Allan. (2012). "Higher-Order Epistemic Attitudes and Intellectual Humility". *Episteme* 9(3): 205-223.

Horowitz, Sophie. (2014). "Epistemic Akrasia". *Noûs* 48(4): 718–744.

Huemer, Michael. (2011). "The Puzzle of Metacoherence". *Philosophy and Phenomenological Research* 82 (1): 1-21.

Kelly, Thomas. (2010). "Peer Disagreement and Higher-Order Evidence". In T. Warfield and R. Feldman (Ed.), *Disagreement*. Oxford University Press. 111–175.

–––– (2014). "Believers as Thermometers". In *The Ethics of Belief*, ed. John Mattheson and Rico Vitz. 301–314.

Korcz, Keith Allen. (2000). "The Causal Doxastic Theory of the Basing Relation". *Canadian Journal of Philosophy* 30(4): 525–550.

Littlejohn, Clayton. (forthcoming). "A Plea For Epistemic Excuses". In Dorsch, F. and Dutant, J., editors, *The New Evil Demon*. Oxford University Press.

Smithies, Declan. (2012). "Moore's Paradox and the Accessibility of Justification". *Philosophy and Phenomenological Research* 85(2): 273–300.

–––– (2016). "Belief and Self-Knowledge: Lessons from Moore's Paradox". *Philosophical Issues* 26: 393–421.

Schechter, Joshua. (2013). "Rational Self-Doubt and the Failure of Closure". *Philosophical Studies* 163: 428–452.

Schoenfield, Miriam. (forthcoming). "An Accuracy Based Approach to Higher-Order Evidence". *Philosophy and Phenomenological Research.* Doi: 10.1111/phpr.12329.

Sliwa, Paulina. & Horowitz, Sophie. (2015). "Respecting All the Evidence". *Philosophical Studies* 172(11): 2835-2858.

Titelbaum, Mike. (2015). "Rationality's Fixed Point (Or: In Defense of Right Reason)." In J. Hawthorne (ed.) *Oxford Studies in Epistemology* 5. Oxford University Press. 253–294.

van Wietmarschen, Han. (2013). "Peer Disagreement, Evidence, and Well-Groundedness". *The Philosophical Review* 122: 395–425.

Weatherson. Brian. (ms.). "Should We Act on Higher-Order Evidence?". Available at <http://brian.weatherson.org/AAP-Talk.pdf>.

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

References

Ballantyne, Nathan. & Coffman, E.J. (2012). "Conciliationism and Uniqueness". *Australasian Journal of Philosophy* 90: 657-670.

Broome, John. (1999). "Normative Requirements". *Ratio* 12: 398–419.

Christensen, David. (2007). "Epistemology and Disagreement: The Good News". *Philosophical Review* 116: 187–217.

Christensen, David. (2009). "Disagreement as Evidence: The Epistemology of Controversy". *Philosophy Compass* 4(5): 756-767.

Christensen, David. (2010a). "Higher-Order Evidence". *Philosophy and Phenomenological Research* 81(1): 185–215.

Christensen, David. (2010b). "Rational Reflection". *Philosophical Perspectives* 24: 121-140.

Christensen, David. (2011). "Disagreement, Question Begging, and Epistemic Self Criticism". *Philosophers Imprint* 11(6): 1-22.

Christensen, David. (2013). "Epistemic Modesty Defended". Jennifer Lackey and David Christensen (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 77–97.

Christensen, David. (Forthcoming). "Disagreement, Drugs, etc.: From Accuracy to Akrasia". *Episteme*.

Cohen, Stewart. (2013). "A Defense of the (almost) Equal Weight View". Jennifer Lackey and David Christensen (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press. 98–120.

Comesaña, Juan. (2015). "Normative Requirements and Contrary-to-Duty Obligations". *The Journal of Philosophy* 112 (11): 600–626.

Comesaña, Juan. (ms.). "On An Argument Against Immediate Justification". Available at <http://comesana.web.arizona.edu/files/on-an-argument-against-immediate-justification.pdf>

Elga, Adam. (2007). "Reflection and Disagreement". *Nous* 41(3): 478–502.

Elga, Adam. (2010). "How to disagree about how to disagree. In R. Feldman and T. A. Warfield (Eds.), Disagreement, Oxford University Press. 175–186.

Elga, Adam. (2013). "The puzzle of the unmarked clock and the new rational reflection principle". *Philosophical Studies* 164: 127–139.

Elga, Adam. (ms.) "Lucky to be Rational".

Enoch, David. (2010). "Not Just a Truthometer: Taking Oneself Seriously (But Not Too Seriously) in Cases of Peer Disagreement". *Mind* 119 (476): 953–997.

Easwaran, Kenny et al. (2016). "Updating on Credences of Others: Disagreements, Agreement and Synergy". *Philosophers' Imprint* 16(11): 1-39.

Feldman, Richard. (2006). "Epistemological Puzzles About Disagreement" in Stephen Hetherington (ed.) *Epistemology Futures* (Oxford University Press, Oxford), 216 -236.

Feldman, Richard. (2007). "Reasonable Religious Disagreement". In L. Antony (ed.), *Philosophers Without Gods*. Oxford University Press, 194-214.

Feldman, Richard. (2009). "Evidentialism, Higher-order Evidence, and Disagreement". *Episteme* 6: 294–312.

Feldman, Richard. (2014). "Evidence of Evidence is Evidence". In *The Ethics of Belief*, ed. John Mattheson and Rico Vitz. 284–99.

Finlay, Stephen. (2010). "What *ought* probably Means, and why you can't detach it". *Synthese* 177: 67–89.

Harman, Gilbert. (1986). Change in View: Principles of Reasoning, MIT Press, Cambridge, MA.

Hazlett, Allan. (2012). "Higher-Order Epistemic Attitudes and Intellectual Humility". *Episteme* 9(3): 205-223.

Horowitz, Sophie. (2014). "Epistemic Akrasia". *Noûs* 48(4): 718–744.

Huemer, Michael. (2011). "The Puzzle of Metacoherence". *Philosophy and Phenomenological Research* 82 (1): 1-21.

Kelly, Thomas. (2005). "The epistemic significance of disagreement". In John Hawthorne and Tamar Gendler (ed), *Oxford studies in epistemology*, volume 1. Oxford University Press, Oxford. 167-197.

Kelly, Thomas. (2010). "Peer Disagreement and Higher-Order Evidence". In T. Warfield and R. Feldman (Ed.), *Disagreement*. Oxford University Press. 111–175.

Kelly, Thomas. (2014). "Believers as Thermometers". In *The Ethics of Belief*, ed. John Mattheson and Rico Vitz. 301–314.

Korcz, Keith Allen. (2000). "The Causal Doxastic Theory of the Basing Relation". *Canadian Journal of Philosophy* 30(4): 525–550.

Kolodny, Niko. & MacFarlane, John. (2010). "Ifs and Oughts". *The Journal of Philosophy* 107(3): 115-143.

Littlejohn, Clayton. (forthcoming). "A Plea For Epistemic Excuses". In Dorsch, F. and Dutant, J., editors, *The New Evil Demon*. Oxford University Press.

Matheson, Jonathan. (2009). Conciliatory Views of Disagreement and Higher-Order Evidence. *Episteme* 6(3): 269–79.

Silk, Alex. (2014). "Why 'Ought' Detaches: Or, Why You Ought to Get With My Friends (If You Want to Be My Lover)". *Philosophers' Imprint* 14(7): 1-16.

Smithies, Declan. (2012). "Moore's Paradox and the Accessibility of Justification". *Philosophy and Phenomenological Research* 85(2): 273–300.

Smithies, Declan. (2016). "Belief and Self-Knowledge: Lessons from Moore's Paradox". *Philosophical Issues* 26: 393–421.

Schechter, Joshua. (2013). "Rational Self-Doubt and the Failure of Closure". *Philosophical Studies* 163: 428–452.

Schoenfield, Miriam. (2015). "A Dilemma for Calibrationism". *Philosophy and Phenomenological Research* 91(2): 425-455.

Schoenfield, Miriam. (Forthcoming). "An Accuracy Based Approach to Higher-Order Evidence". *Philosophy and Phenomenological Research.* Doi: 10.1111/phpr.12329.

Sliwa, Paulina. & Horowitz, Sophie. (2015). "Respecting All the Evidence". *Philosophical Studies* 172(11): 2835-2858.

Smities, Declan. (2012). "Moore's Paradox and the Accessibility of Justification". *Philosophy*

and Phenomenological Research 85(2): 273–300.

Titelbaum, Mike. (2015). "Rationality's Fixed Point (Or: In Defense of Right Reason)." In J. Hawthorne (ed.) *Oxford Studies in Epistemology* 5. Oxford University Press. 253–294.

van Wietmarschen, Han. (2013). "Peer Disagreement, Evidence, and Well-Groundedness". *The Philosophical Review* 122: 395–425.

Way, Jonathan. & Whiting, Daniel. (2015). "If You Justifiably Believe That You Ought to φ, You Ought to φ". *Philosophical Studies*, doi:10.1007/s11098-015-0582-2.

Way, Jonathan. (2010). "Defending the wide-scope approach to instrumental reason". *Philosophical Studies* 147: 213–233.

Weatherson, Brian. (ms.). "Do Judgments Screen Evidence?" available at <http://brian.weatherson.org/JSE.pdf.>

Weatherson, Brian. (ms.). "Should We Act on Higher-Order Evidence?". Available at <http://brian.weatherson.org/AAP-Talk.pdf>.

White, Roger. (2005). "Epistemic Permissiveness". In J. Hawthorne (ed.), *Philosophical Perspectives* 19, Epistemology, Malden, MA: Blackwell Publishing, 445–59.

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