Philosophers' Imprint

VOLUME 18, NO. 13
JULY 2018

Anticipating Failure and Avoiding It

Robert Steel

1. Introduction

When we try to figure out what to believe, having evidence is good and having more evidence is even better. We like the following sequence of events: I get some new evidence, and then, on that basis, I update my beliefs. Call this The Process.

> The Process: I hold some beliefs P1, P2, etc. on the basis of some evidence E. I acquire some additional evidence, E*, and then on that basis adjust my previous beliefs so as to arrive at some new P1*, P2*, etc.

I take it to be a basic datum in epistemology that The Process is generally good. It improves our beliefs. That's why we go to such great trouble looking for more evidence, and why, when we are faced with decisions of great consequence, it is especially important that they be made on a maximal evidential basis.

I said above that The Process improves our beliefs. What is it to improve our beliefs? One natural standard is matching; beliefs improve when they match the world better. When we work within the full belief framework, this can be cashed out in terms of truth and falsity; we say our full beliefs match the world when they're true. But when we focus on partial beliefs rather than full beliefs, truth can no longer play that role directly—on the usual understanding, a degree of confidence, or credence, cannot, strictly speaking, be true or false. So instead we speak of the analogue, graded notion of "accuracy"; we say a partial belief matches the world *better* when it's *more accurate*.[1]

The platitudes above about The Process generally improving our beliefs are true on either framework; getting and responding to evidence tends to give us true full beliefs and accurate partial ones. But since this paper is interested particularly in rational degrees of belief, I will focus on the second version. So: The Process is generally good, where by that I mean that going through The Process generally increases the accuracy of one's beliefs.

1. To the extent that the differences among them are relevant, I discuss the particular ways in which accuracy can be quantitatively measured in section 9.

But even if The Process is generally good, it need not be good in every instance. Sometimes it fails, where by failure I just mean the converse of what I meant by success — namely, that it sometimes issues in beliefs less accurate than those with which it began.

In this paper I begin by looking at two ways The Process can fail. The first: I receive misleading evidence. I faithfully update my beliefs in light of that evidence, but because it is misleading it points me astray. The second: I receive non-misleading evidence. But because I manifest some irrationality, I update my beliefs in ways contrary to the truth that evidence in fact suggests. The evidence points in the right direction, but I go astray anyway.[2]

What should I do if I anticipate that some evidence I'm about to get is such that, in taking account of it, I'm likely to go wrong in one of these two ways — that, one way or another, responding to it by way of The Process will degrade my accuracy? Answer: If I anticipate that taking account of it will lead me astray, I ought to make sure I don't take account of it. Rather, I ought to ensure I avoid or ignore it. To do otherwise would involve taking what I anticipate to be a bad epistemic deal.[3]

In this paper, I show that this seemingly anodyne answer has surprisingly substantial consequences for the current debate on peer disagreement. Specifically, I show how it can be used to militate in favor of a "conciliatory" view. Such views take the discovery of disagreement with one's peers to have serious corrosive effects on one's rational confidence — the version of this doctrine I espouse is roughly Elga's, and the details will be spelled out in section 3.

How does the claim that we should avoid taking bad epistemic deals motivate conciliation? It is common among non-conciliatory views to posit some evidence, available in cases of disagreement, that goes unrecognized by the conciliationist and which rationalizes whatever non-conciliatory degrees of confidence they take to be appropriate. But on closer examination it becomes clear that this evidence is such that we can anticipate, in advance, that in responding to it The Process will fail in one of those two ways: it's built into the nature of this evidence that trying to respond to it leads to failure. And if that's right, then, as above, we should not try to respond to it — rather, we should ignore it. Once it is ignored, non-conciliatory views lose their rationale.

The upshot is that conciliatory doctrines about what one ought to believe turn out to be surprisingly autonomous from questions of what the relevant evidence supports.[4] We can allow that the conciliationist is indeed ignoring all sorts of genuine evidential features of the situation, just as their critics allege: we are still left with a surprising and powerful argument that ignoring those genuine evidential features is exactly what they should be doing.

Here's the plan: I begin in section 2 by discussing some ordinary examples in which we expect The Process to fail, with an eye to motivating the idea that in such cases we ought to avoid basing our beliefs in the problematic evidence. I then move on in section 3 to introduce my understanding of what it is to be a peer, as well as the content of the conciliatory doctrine I espouse. Sections 4 through 7 introduce alternate non-conciliatory views and explain how they construe the evidence in disagreement; sections 8 and 9 show how the evidence, so construed, is such that we expect ahead of time that it will cause The Process to fail. With these pieces assembled, I am then in a preliminary

2. This list of ways to fail is not intended to be exhaustive.

3. The focus on what one anticipates may prime the reader to expect an argument grounded in some reflection principle or other. I explain why that's not the right way to gloss my argument after I've finished giving it, in section 15.

4. I am indebted to Miriam Schoenfield's (2012) for helping me see that there are sensible reasons to multiply our epistemic concepts and separate out what one's evidence supports from what one should believe. See also van Wietmarschen (2013), which argues in a very different way for a conclusion with a similar ring to it: that conciliation is a bad theory of what the evidence supports, but that it nonetheless may be a good theory of well-grounded belief.

   Of course, another option is to insist on the close connection between evidence and what one should believe. That is not my preferred route, but the argument of this paper does not hinge on that choice. If one insists on an inseparable connection between evidence and rational belief, then read the argument of the paper as: Non-conciliatory views entail the existence of evidence you should ignore, of which there can be none, and so they must have gone wrong somehow.

position to draw my desired conciliatory conclusion: since we ought to ignore pieces of evidence that we anticipate will cause The Process to fail, and the evidence posited by non-conciliatory views has this feature, it follows that we ought to ignore the evidence posited by non-conciliatory views — and, in so doing, embrace conciliation.

Section 10 outlines an objection, to the effect that although the evidence posited by non-conciliatory views is such that we expect it to cause The Process to fail *in general*, in each particular instance we are in a position to endorse our responses. Sections 11 through 14 answer this objection.

Finally, section 15 steps back and situates the argument relative to some more familiar epistemic arguments and claims, before I go on in section 16 to conclude.

2. Anticipating Failure

The Process is generally good. But what precisely do we mean when we say that? We cannot mean that, in each instance in which The Process occurs, the person undergoing it ends up with more accurate beliefs than they had when they started. Sometimes things go wrong, and they instead end up at a worse place than they started. One way this can happen is when misleading evidence leads them astray. Like everyone else, I started this morning believing the world's nuclear missiles were safely in their silos. But then I saw a report on the local news: they have been launched! So I believed. But in fact the news was being pranked, and there was no launch. In this case, the report was misleading.

Schematically,

> Failure Due to Misleading Evidence: I hold some beliefs P1, P2, etc. on the basis of some evidence E. I acquire some additional evidence, E*, and then on that basis adjust my previous beliefs so as to arrive at some new P1*, P2*, etc., which are supported by E*. However, E* is misleading, and P1*, P2*, etc. are less accurate than P1, P2, etc. were.

How do we understand the general goodness of The Process such that it's consistent with its potential to fail due to misleading evidence? It's normal to want to say something like "Misleading evidence is unusual" or "*Typically* evidence is not misleading" or the like. How to cash out these thoughts in a detailed way is not obvious. However, as the interest here is not in a blanket skeptical question, I will take myself to be entitled to the following minimal anti-skeptical conclusion: the rational default is to suppose evidence isn't misleading. Until one has some reason to think otherwise, one ought to treat one's evidence as non-misleading, and, concomitantly, one ought to apportion one's beliefs as it suggests.

This already goes some way toward reconciling the general goodness of The Process with the possibility of failure due to misleading evidence. If the default is to regard evidence as non-misleading (even though sometimes it is), then the default is similarly to regard instances of The Process as improving accuracy (even though sometimes they don't). And if this is the rational default attitude, we can also rationalize the broad data mentioned in the introduction, namely our general interest in acquiring more evidence and our conviction that acquiring more evidence is especially vital when something important hangs in the balance.

Getting misleading evidence is one way The Process can go wrong. But here is another: Sometimes we acquire perfectly non-misleading evidence, yet we respond to it irrationally. As it were, we acquire evidence that points in the right direction, but we nonetheless go in the wrong one.

For instance: suppose I am superstitious. I begin my day by thinking that this workday will be like any other. But on the way to work I notice foreboding clouds on the horizon. I think to myself, "This is a bad sign." I become convinced that I will be fired. Still, the day passes without event.

We can stipulate that the presence of ominous clouds on the horizon was not misleading evidence for the false proposition that I would be fired. This because it was not evidence for the proposition that

I would be fired at all; clouds are evidence of rain, etc., not general heralds of misfortune. It was only by virtue of my irrational superstitions that I came to be convinced on the basis of the perfectly innocent clouds that I would later be jobless.

Schematically,

> Failure Due to Irrationality: I hold some beliefs P1, P2, etc. on the basis of some evidence E. I acquire some additional evidence, E*, and then on that basis adjust my previous beliefs so as to arrive at some new P1*, P2*, etc. However, despite the additional evidence E* being non-misleading, my response to it is irrational; E* does not actually support P1*, P2*, etc. And, as it so happens, P1*, P2*, etc. are less accurate than P1, P2, etc. were.

This is another way The Process can leave us worse off than we started.

Again, as a minimal anti-skeptical commitment, it is natural to say that, in some sense, failure due to irrationality is the unusual case. Even if we normally fail to be maximally rational, we usually get close enough that The Process still leaves us at least better off than we were when we started. And this fact seems like it ought to have something to do with why the mere possibility of failure due to irrationality is compatible with a default presumption in favor of The Process. These remarks are cursory and, even in this vague form, controversial; it would be the task of a different paper to give them in a more perspicuous form. For present purposes, just allow that there is some way, which we leave obscure, to describe the interaction of merely possible failure due to irrationality with the general case such that the general goodness of The Process can be affirmed.

So there are some default presumptions to the effect that for any arbitrary piece of evidence E, if we were to attempt to assimilate it by way of The Process, then we would thereby come to have more accurate beliefs; the default is against assuming that E is misleading or that our response will be irrational in an accuracy-destroying way.

What I want to ask now is: What happens when that default assumption is disrupted? That is to say: What happens when we have strong evidence, for some piece of evidence E, that it is misleading, or that our evaluation of it would be irrational?

Let's look at some cases. So, for instance, imagine that you're me in one of the above two scenarios. But this time, rather than being caught unaware, you hear from someone beforehand: "The radio station's getting pranked today. Someone's going to tell them that there's a nuclear war on!" Or, alternately: "That stuff you're always doing with clouds doesn't make a whole lot of sense. You know clouds are just vapor, right?" If these comments are sufficiently persuasive, you may then come to agree.

If you are so convinced, then when you hear the radio report, or you see the ominous clouds, you will ignore them. If one becomes convinced beforehand that some piece of evidence will be misleading, or that the response it provokes will be irrational, then that can in turn rationally change the way that one responds to that evidence. So long as one has the power to ignore, there is no threat that one will be somehow forced to make what one foresees as a bad epistemic move. We may say: So long as one has the power to ignore, one never really acquires evidence that one *will*, by way of some piece of evidence E, update one's beliefs to become less accurate. At most one can have evidence that one *would have* updated one's beliefs with E in such a way as to become less accurate — *would have*, had one not known this very fact. But since one does know this fact, one instead just ignores E.

Unfortunately, not all cases are so easy. Sometimes we may rationally become convinced beforehand not only that some evidence E will be misleading, or that some response to E will be irrational, but that we cannot avoid that result by ignoring E. This because we cannot ignore it — if we acquire it, we simply *will* respond in a way so as to degrade our accuracy. One example is information about e. g. race and gender. Suppose the evidence in question contains facts about someone's race or gender. And furthermore, suppose that you believe, as many now do, that social scientists have provided definitive evidence

that most people cannot simply ignore facts about race and gender (even if you do not in fact believe this, imagine that you do). And so suppose that we take it as a given that race and gender simply *do* influence all sorts of evidential assessments whether we'd like to ignore them or not.

What is to be done then? If we cannot ignore the evidence, and we know ahead of time that, when encountered, it will have deleterious effect, we need to do our best not to get it; to forget it when we get it; to prevent ourselves from being in a position to act on it after we get it; and so on. That sounds odd in the abstract, but the mechanisms by which it can be carried out are not so unfamiliar. Suppose that I am hiring a new member for my chamber orchestra. I may choose to listen to the auditions from behind a screen that obscures the tryouts — my purpose in doing so is to prevent myself from learning the races and genders of the auditioners. I believe that implicit biases can cause people to misevaluate musical talent, and given my belief I take utterly sensible steps to prevent my assessment from being so corrupted. In short, I make sure not to learn certain facts about the people I'm hiring.[5]

That's one kind of case, one that is often glossed in terms of *unconscious* influences on our beliefs that may act so as to bend them away from truth and toward stereotype. But ordinary examples needn't be unconscious. Rather, we can have strong evidence that our explicit, fully articulated, and fully conscious judgments in response to some future piece of evidence will involve a failure due to irrationality — even knowing full well and ahead of time that this is so.

To illustrate: suppose I have spent many years in an abusive relationship. I have finally found an opportunity to walk out, but have left some prized possessions behind at my former boyfriend's place. I want to get them back, but I'm afraid that if I return I'll see him. And I'm afraid that if I do he'll talk me into taking him back. I am extremely sure that leaving him is the right decision, but at the same time I know that he has, over the years, developed power over me, and that he understands how to break down my resistance.

Take the evidence here to be the content of the speech my ex-boyfriend is planning to give me. Here I have strong evidence that, when it comes to this particular piece of evidence, if I encounter it, then The Process will fail due to irrationality. I have evidence that even though nothing he says will present an actual reason for me to return, I will nonetheless be convinced by his speech to return. And this even if I go into the situation with a steely resolve, if I tell myself beforehand not to be fooled by him again, and so on, and even though the reasoning that I anticipate being faulty is of the fully explicit sort that we ordinarily classify as under our direct control. So again, what I ought to do is make sure I do not give my ex a chance to talk to me. I ought to send a friend to get my things, or perhaps I ought to give up on them entirely.[6]

5. The adoption of this screened audition process in major orchestras has coincided with a substantial growth in their proportion of female musicians. For an analysis suggesting that a significant portion of this growth is indeed attributable to the adoption of the screened auditions, see Goldin and Rouse (2000).

6. There is something of a parallel here between these claims and some made in the literature on overall practical obligation. I say: When you anticipate that you *will* respond irrationally to evidence, you should avoid or ignore the evidence. Actualists about practical obligation say: When you anticipate that you *will* act badly in a choice situation, you should avoid the choice. These claims seem close, and similar intuitive cases can be made to tell in favor of both. Possibilists, though, say: What you *should* do is not avoid the choice — rather, you should encounter the choice and act well (even if this is not what you anticipate will happen, it is all the same what you *should* do). Given that my position looks like an epistemic analogue of actualism, and actualism opposes possibilism, does that mean that possibilists ought to reject my position? Although there is a parallel, I'd suggest the two lines can be pulled apart. Here is one disanalogy: in the cases which actualists and possibilists argue over, the subject knows what the practically obligatory choice is and the problem is that they anticipate they may be too weak-willed to select it. By contrast, in the cases I discuss, the subject has no problem with epistemic continence. They anticipate believing that which they take to be well-supported; they just think they will likely be wrong about what that is. This difference seems, at least to me, to be significant enough. So, there is at least some room to doubt that all salient facts about practical rationality port over to the theoretical case, and so at least room to think that my argument here does not require taking a controversial stand on the issue. Thanks to an anonymous reviewer for calling this parallel to my attention.

The Process is generally good, and surely this is part of the rationale for our general desire to base our beliefs in the most evidence possible. But focusing too hard on the general goodness of The Process can lead us to mistakenly make the absolute claim that for *any* piece of evidence, we should *always* want to acquire and take account of it. Attention to fairly ordinary cases can help dispel this thought. More evidence may usually be better, sure, but all the same there are cases where our attitude toward a potential body of new evidence is decidedly negative. In these cases, we may well recognize the existence of some evidence while nonetheless wanting very much not to acquire it; if we acquire it, we may not want to consider it; if we consider it, we may not want to base our beliefs off it; and so on. I take it that the cases in this section fit that pattern.

3. Peer Disagreement and the Conciliatory View

Perhaps we generally respond irrationally to certain socially loaded categories, like race and gender. And one's absolutely catastrophic history of decision-making with regards to one's ex-boyfriend may present strong evidence that one will respond irrationally to his entreaties. Those are two fairly particular instances where we may entertain rational expectations of future failure. Are there any more general classes that we can fruitfully analyze?

Yes. In particular, in this paper I will be focused on the philosophical treatment of disagreement. Broadly: when we learn that those around us whom we take to be reasonable and well-informed have come to differing conclusions, that may create a rational expectation of failure.

More specifically: in treating disagreement I will focus on one extremely special case — the case where one finds oneself disagreeing with someone whom one takes to be, in a specific sense, an "epistemic peer". I make this choice because there already exists a literature on this case, and because its artificiality makes it tractable. I think the verdict I reach will generalize broadly; others may disagree. Regardless, before we can worry about the proper generalization, we need to get the case itself right — and given the many divergent treatments it's received, getting straight on that is already a big enough task for this paper.

The special case is that one holds the following conditional credence: For some person A, and some proposition P, conditional on one's disagreeing with A over P, one's credence in P is ½.

An example where such a credence would arise:

> *Two Bright Students:* Misty and Ash are good friends. They are also both clearly the best students in their class. Their averages have hovered around the same stellar number throughout the course so far, and they have traded the top class rank back and forth. In the past, when they have disagreed over the answer to a problem, they have each been right an equal portion of the time. Their final exam, rapidly approaching, consists in a single true-or-false question. This question will be neither so easy that they can just "see" the answer, nor so difficult that they are just hopeless at figuring it out: it lies in the broad range of questions that they have been tackling with roughly equal success over the course of the semester.

I claim that in the above case, it is rationally required for Misty to think of Ash that if, on the final exam, Ash disagrees with her over whether the answer to the final question is true, then it's 50 / 50 which of them is right — and, by extension, it's 50 / 50 whether the proposition in question is true. So the story above is one on which Misty is rationally required to hold of Ash the special sort of belief I'm interested in.[7]

7. It's worth noting that I identify conciliationism, the position I am concerned to defend, with a particular way of handling this special belief. But there are people who can be conciliationists in that sense while still retaining what are intuitively "egoistic" and/or "nonconformist" overall views of disagreement — as, for instance, if they thought it *a priori* almost impossible to get into the situation of regarding someone else as your peer in the specified sense. See for instance Schafer (2015) and, as noted in ftn. 12, possibly Wedgwood (2010). I think those views can be ruled out, but doing so requires a separate argument from the one given here. It is also worth noting that this is not the

Now suppose:

> *A Frustrating Final:* Misty and Ash take the final exam. As they walk out, Ash turns to Misty and says, "I got F." Misty replies, "Uh-oh, I got T."

Having learned of Ash's disagreement, what should Misty think?

The answer may look trivial: She should think it's 50 / 50! If Misty had the conditional credence described, doesn't it just follow that she should update according to it? It does not. Even supposing that conditionalization is the correct general way to update one's beliefs, it's still the case that one ought to conditionalize on the *strongest* evidential proposition one has learned. Misty's earlier belief was conditional on her and Ash disagreeing over the final question. Over the course of the story told, is "the final exam happened and Ash and I disagreed over the final question" the *strongest* relevant thing she learned?

Call this fact "the mere fact of disagreement". The question then becomes whether, during the course of taking the final and then comparing answers with Ash, Misty learned anything epistemically relevant over and above the mere fact of disagreement. If the answer is that she learned something important, then Misty may need to change her mind. She may then no longer think it's 50 / 50.

Of course, Misty may notice that Ash is visibly drunk when he's taking the test, or alternately that he has cheated and smuggled in the answers, or whatever, and thereby come to be very sure that he's either right or wrong. That would be fine so far as it goes, but it's not what we're interested in. Rather, the question is about whether there's anything learned in cases of disagreement *per se* over and above the mere fact of disagreement.

only way of defining what it is to be an "epistemic peer". That's fine; my thesis should be understood just as the claim that epistemic peerhood *understood in this way* requires a particular response. I take no stand on what other notions of peerhood require. See e.g. Lam's (2011) and (2013) for examinations of how varying the understanding of peerhood in play can correspondingly vary the performance of different disagreement policies, and see ftn. 11 for further discussion of Lam's work.

So there is a substantive question here. I will argue that the answer to that substantive question is the one that you might have thought was trivial: Misty should still think it's 50 / 50. If she learns anything over and above the mere fact of disagreement, it's not the sort of thing that can rationalize a change in confidence. Call this view the conciliatory view.[8]

4. Independence and Extra Weight

What might Misty learn in the course of the disagreement, over and above the mere fact of disagreement? Well, she sees the question itself. She reasons through it. She arrives at her answer. Beforehand, she was agnostic about who would be right if Ash disagreed with her on the final question, whatever that was. But now she sees that Ash is disagreeing with her not just on a question, but on *this* question, and that he is doing so by getting *that* answer.

How does this change things? Suppose Misty reasons as follows:

> D: I got T on the final question. And T is the answer to the final question. Ash got F. Since T is the right answer, Ash's answer is wrong. Since Ash got the wrong answer, he must have reasoned incorrectly in this case. Since I can now see that he reasoned incorrectly, our situation is no longer symmetric. I no longer take him to be my peer when it comes to this specific question, and if I need take any account of his opinion at all, it need not be as much as I antecedently would have thought.

In reasoning in this way, Misty leverages the very facts under dispute in the disagreement while considering how she ought to respond to it. If that's legitimate, then it will never be the case that thinking it was 50 / 50 in advance of the disagreement requires thinking it continues to be 50 / 50 afterward. Rather, one will always discover

8. This view is popular in the literature, and leading defenders include e.g. Christensen (2007), (2009), (2011), and Elga (2007), (2010a).

something new, something capable of disturbing one's original assessment: namely, the answer one got. And so conciliationism would be false.

The reasoning in D, however, can seem problematic precisely by virtue of its appeal to the very facts under dispute. To block such reasoning, conciliationists have proposed the following principle:

> INDEPENDENCE: In evaluating the epistemic credentials of another person's belief about P, in order to determine how (if at all) to modify one's own belief about P, one should do so in a way that is independent of the reasoning behind one's own initial belief about P.[9]

If Independence is true, then reasoning as in D is illegitimate. And if reasoning as in D is blocked, then it is plausible that there is nothing relevant that Misty learns over the course of the disagreement to disrupt her initial assessment. In other words, conciliationism looks to follow. Both ways, conciliationism seems to stand or fall with Independence.

Ought we to adopt Independence, and by doing so block reasoning like that in D? One reason to think that we should is that the reasoning in D looks dogmatic, and as such epistemically unimpressive. This negative first impression can be made more precise. Suppose we interpret reasoning as in D as licensing Misty to, in the course of disagreement, give preferential treatment to her own convictions over Ash's merely on the basis that they are her own — this is at least one way to fill out the dogmatic undercurrent to D. Since this interpretation of D involves giving one's own view extra weight, call this view *the extra weight view*.

Elga has argued against the extra weight view by pointing out that it appears to allow Misty to "bootstrap" her way into undeserved confidence in her own abilities; since such bootstrapping is epistemically abhorrent, so too must be views which permit it.

How does the bootstrapping objection proceed? Elga offers the following *reductio*:

> Suppose it really were permissible to give one's own view extra weight. If that were so, then after disagreeing with a peer, one could be rationally confident that one was right; but if *that* were so, then one could also become rationally confident in the propositions one's having been right entails, namely that one has a better track-record than one's friend, and hence that one is more reliable after all — and all this would be possible in advance of actually receiving any independent confirmation of one's rightness. If this were really rational, that must be because disagreement with a peer is evidence that one is the better judge. But that's absurd. The mere fact of disagreement, absent any independent confirmation that one is right, is no evidence at all that one is the better judge. Hence, the extra weight view is false.[10]

This line of objection is raised against views which would allot oneself extra weight, but it applies just as much to views which would allot one's interlocutor extra weight: there the absurd conclusion is just the reverse of the one derived before, namely that one can take the mere fact of disagreement to be evidence that one's interlocutor is more reliable. The point here is just that the existence of disagreement on its own is no evidence at all either way. Therefore, Elga's bootstrapping objection aims to rule out any deviation from the conciliatory view whatsoever.

As it stands, though, this argument does not do much to motivate Independence. The bootstrapping argument assumes that the mere fact of disagreement is the only epistemically relevant thing learned

9. This principle is proposed in e.g. Christensen (2007, p.16–17) and Christensen (2011, p.1–2). Independence, as formulated here, does *not* block off making reference to the existence or even the properties of one's reasoning — again, provided that those properties are picked out in a way that does not presuppose the correctness of that reasoning.

10. Elga (2007, p.12–15).

and then, on that basis, constructs a dilemma where the choices are conciliation or extra weight. And it's true that, if that were the dilemma, then conciliation would be thereby established as the only sensible option. But the assumption which generates the dilemma, namely that the mere fact of disagreement is the only epistemically relevant thing learned, is something that itself needs to be demonstrated.

Indeed, non-conciliatory theorists have responded to the bootstrapping argument by allowing that it is fatal to the extra weight view — where that view is the one that takes the mere fact of disagreement to be evidence — but then carefully distinguishing their own views. They have done so by giving an account of the evidence they take to arise over the course of disagreement, over and above the mere fact of disagreement, and describing it in such a way that it seems to be the sort of thing that could rationalize a change in view. These replies can successfully defuse the objection as it is phrased above: there it takes the form of accusing non-conciliatory views of taking the mere fact of disagreement to be evidence when it is not, in fact, really evidence. When non-conciliatory theorists give an account of the more robust evidence that they take to arise in disagreement such that it looks like it is the right sort of thing to do the work required, then they adequately respond to the objection.

I will argue, however, that there is a closely related problem which has not been the subject of so much attention. Elga's objection can be reformulated not in terms of the evidence that there is, but in terms of how one knows one will respond to it. Even though non-conciliatory views can give a richer description of the evidence at hand than the extra weight view does, they nonetheless bear a problematic relationship to it: namely, we can ascertain ahead of time that if one tries to follow a non-conciliatory view, one will in fact reach the same results that one would have reached by actually following the extra weight view. But following the extra weight view looks like a bad deal from the perspective of accuracy — and since accuracy is a matter of what you believe, not why you believe it, if the extra weight view looks bad from the perspective of accuracy, then so too do all the views the attempted following of which would yield the same results.[11]

Before I give this argument in greater depth, I need to introduce the non-conciliatory views I am to criticize. I will canvass three major alternatives. First I'll describe Enoch's steadfast view.[12] Then I'll describe Kelly's total evidence view.[13] Finally, I'll describe the right reasons view; since this last is, in all relevant structural features, just a variant of the

11. My approach here bears some notable similarities to the line advanced by Barry Lam in his (2011) and (2013). He seeks to inform the normative debate over peer disagreement by answering closely related non-normative questions about how different views will perform under application; I try to inform the normative debate over peer disagreement by answering non-normative questions about how agents expect their views to perform under application. I take it that the most significant difference in our approaches is that we focus on different understandings of what it is to be a peer — I understand peerhood in terms of the holding of a conditional credence of the type outlined in section 3, whereas he understands peerhood in terms of agents equally well satisfying a measure of epistemic success (he considers multiple such measures). These different operative understandings of peerhood lead to significant downstream differences in our conclusions. It is interesting to sort out which of these different conclusions represent actual disagreements and which simply answer different questions; my sense is that there is much of the latter, though I can't fully trace that out here.

12. Enoch (2010). By contrast to Enoch, it is unclear to me exactly how to categorize Wedgwood's view with respect to conciliation. In his (2010, p. 237) he seems to endorse Independence, while claiming that it will be hardly ever applicable, because one hardly ever ought to have the conditional credence that characterizes peerhood beliefs as I've defined them. Yet later (2010, p. 243) he seems to endorse reasoning relevantly like D. I suspect this might be because he takes it to be the case that one typically lacks any relevant credence about one's peer's reliability prior to some disagreements, and hence reasoning like D need not conflict with Independence. This would, in my locution, make him a conciliationist, albeit one who does not think conciliationism is a particularly interesting or widely applicable thesis.

13. Kelly (2010) and (2013). One significant view that I do not explicitly discuss is Lackey's "justificationist" view. I take the justificationist view to be, in the respects relevant to the argument of this paper, the same as the total evidence view: the justificationist's reliance on one's *actual* initial degree of justification, as well as on the *actual* propriety of one's mental goings-on as encoded in "personal information", will together have the same effect as the total evidence view's reliance on the *actual* disposition of the non-psychological evidence. So, I take my criticisms of the total evidence view to apply *mutatis mutandis* to the justificationist view. Lackey (2010a), (2010b).

total evidence view, my description will be quite brief. The point will be to get onto the table their conceptions of the relevant evidence that arises over the course of disagreement. Once that's done, I can go on to argue that, as so specified, that evidence bears a troubling relationship to accuracy.

5. The Steadfast View

First, what is Enoch's position? He argues as follows: The only way in which it is possible to form judgments about who is a peer and who is not is to look to someone's track record: how reliable has she been on this subject in the past? Furthermore, the only way to do that is to compare her history of judgments to one's own and see how well they match — we have nowhere else to start, when judging the reliability of our peers, than from our own views. But if we methodologically take ourselves to be right in our evaluation of a putative peer's past reliability, then it would be arbitrary not to do the same with our present conflict. Thus, Enoch advises, we ought to respond to peer disagreement by, among other things, demoting our putative peer to some degree; her new track record, this judgment included, is worse than her old one. So Independence is false, and the reasoning in D is legitimate.

But doesn't this involve treating the mere fact of disagreement as evidence that one is a better judge that one's peer? It does not. Enoch is careful in differentiating his view from the extra weight view: he insists that the grounds on which we demote our peer is *not* the mere fact of disagreement, but rather it is the deterioration of her track record. And the deterioration of her track record *is* good evidence that she is not a good judge.[14]

To illustrate: suppose I take you to be my peer on matters p-related. I judge p. I then discover that you have judged not-p. On Enoch's view it is appropriate for me to demote you, not on the grounds *that I judged p* and *that you judged not-p*, but rather on the grounds *that p* and *that you judged not-p*. Of course, what makes it the case that it's appropriate for me to demote you on the grounds *that p* is that I have judged p, but that does not make judging p itself my grounds. Our beliefs are, and necessarily must be, transparent insofar as we act from their contents *directly*, rather than from hedging propositions like 'I judge p.'[15]

Since I am acting directly from the content of my judgment, rather than from the fact that I so judged, my judgment is based in your deteriorating track record itself rather than merely the fact that I so judged. There is some relevant evidence here, the evidence of track records, and that distinguishes the steadfast view from the extra weight view.

6. The Total Evidence View

In cases of peer disagreement, Independence screens off the particular belief in question — thus prohibiting 'p, therefore you're wrong about ~p' — but that's not all it screens off. It also screens off the reasoning used to arrive at p. There are obvious reasons for this. Imagine p was screened off but the reasoning supporting it was not. Then one could simply re-conclude p from that reasoning, and then proceed as in D. But Independence is designed to block D. So Independence must screen off that supporting reasoning as well.[16]

This, however, opens up Independence to the charge that it throws out evidence. After all, if I was correct in my reasoning from my evidence to p, then that evidence really does support p. I should not ignore it when coming to my final opinion. Conciliation falsely treats the case as if the only evidence available in the wake of a disagreement with a peer were our differing beliefs. But there is more evidence than that — there is the evidence on which we based our differing judgments in the first place.

So, for instance, Kelly imagines a case of peer disagreement where we both initially form our views on the basis of some evidence E. After

14. Enoch (2010, p. 981–986).

15. Wedgwood (2010, esp. 242–243) both makes this same point and takes it to be central to the epistemology of disagreement, which is why it is tempting to classify him as steadfast despite his aforementioned seeming endorsement of Independence (in ftn. 12).

16. Christensen makes the point in (2011, p.18) when introducing a problem about how to define the scope of Independence.

we consult with each other and discover our differing conclusions, our new total evidence includes both the original evidence E and the fact that we reached contrary responses; call this new pool of evidence E*. He then puts the point as follows: "Notice that, on the Equal Weight View, the bearing of E on H turns out to be completely irrelevant to the bearing of E* on H. In effect, what it is reasonable for you and I to believe about H … supervenes on how you and I respond to E …. E gets completely swamped by purely psychological facts about what you and I believe. … But why should the normative significance of E completely vanish in this way?"[17] The original evidence, E, is still available to us, and so it should not vanish.

But what of that original motivation for Independence — ruling out D? Would Kelly assent to 'p; you think ~p; therefore you are a worse judge of p than I'? The answer, it turns out, is 'It depends.' If, after adding the fact that you think ~p to my total evidence, it still supports p on the balance, then I am allowed to so proceed. If it does not, I am not.[18] The philosophical assumption that there is any general epistemic rule that can be decided beforehand is false. Of course, deciding what the total evidence supports in any particular case is hard, and certainly more difficult than simply applying a formal rule — but "such are the burdens of judgment".[19]

Kelly's view, then, is also distinct from the extra weight view. Whereas the extra weight view tells you to assign extra weight to your own answer as such, Kelly's view tells you to attend to the total weight of the evidence; if the evidence really does favor the answer you initially came to, then you should respond to it by retaining a higher credence in your own answer. But if it favors something else, you ought to believe that other thing. What is important here is just the disposition of the evidence, and nothing in that description makes reference to your view as such.

17. Kelly (2010, p. 124).

18. Kelly (2013, p. 43–45).

19. Kelly (2013, p. 52).

7. The Right Reasons View

The total evidence view says: Conciliatory views overlook the non-psychological evidence. When Ash and Misty disagree about the answer to the final exam, the facts that each holds the view that they do may constitute some symmetrically balanced evidence in favor of each of their answers. But there is the further non-psychological evidence to consider, and it may well break the symmetry between them.

The right reasons view expresses the same core idea in more extreme form. Right reasons theorists hold that not only is there some relevant non-psychological evidence, but the psychological evidence never matters as such. Ash and Misty should both just believe whatever the non-psychological evidence supports, and that remains the same no matter how many people's dissenting voices they are exposed to.[20]

8. Non-Conciliatory Views Functionally Entail Extra Weight

So: all of these non-conciliatory views posit something epistemically relevant which comes into Misty's possession over the course of the disagreement. That may be a reason based in a changing track record, or it may be some combination of the evidence on the final and then the discovery of Ash's contrary verdict, or it may be entirely the evidence on the final. Call this epistemically relevant stuff, whatever it is, E. The non-conciliatory views say: E is the sort of thing that rationalizes deviating from the 50 / 50 verdict with which Misty began. So non-conciliatory views have a story about why both Independence and the conciliatory views that rely on it are false. Independence would exclude, and conciliatory views ignore, this E.

20. Right reasons theorists may allow that there are other propositions for which the psychological evidence provided by disagreement matters — for instance, that all the students answered T may be evidence that it's rational to conclude T. The distinct commitment, though, is that it is nonetheless not (in this case) evidence that T. For more on the possibility 'level splitting' see e. g. Weatherson (2013), (ms), and some passages in Kelly's earlier (2005) view. For a defense of right reasons without level splitting see Titelbaum (2015).

But forget for a moment what E rationalizes. That is to say: forget what Misty *should* do with it. It's worth asking instead what Misty *will* do with it, in the course of trying to do what she should. And on this, proponents of the non-conciliatory views we have considered agree. Misty will respond to E by becoming more confident that she, rather than Ash, has had the better of the disagreement.

We can introduce the following bit of terminology: let's say that epistemic view A *functionally entails* epistemic view B iff someone who attempts to act in conformance with A will, in fact, wind up arriving at all the same attitudes as someone who was actually acting in conformance with B. What non-conciliatory theorists concede — and it is hard to see how they could not — is that their views all functionally entail the extra weight view.

This is easiest to see with respect to the steadfast view. On the steadfast view, both Misty and Ash receive an E such that each of them may become rationally confident in their own correctness on its basis. But this is just the same result that would be arrived at by parties following the extra weight view. As such, the steadfast view functionally entails the equal weight view.

Enoch acknowledges this but is untroubled. He says one can foresee, on his view, that one will arrive at all the same verdicts as the person who favored their own view with extra weight, but nonetheless the person following his view does not act under the intention of giving herself extra weight as such, and this difference between foresight and intention is epistemically relevant.[21] So he acknowledges the functional entailment while disputing its significance. But that's fine; the point for now is just that there is such a functional entailment.

The total evidence view is not symmetrical in the way that the steadfast view is. On the total evidence view, Misty and Ash receive a single E, and that E is such that it is tilted toward the person who was in fact better responding to the evidence when they formed their initial judgment. So both of them ought to arrive at the same judgment, namely one tilted toward the person who was in fact better responding to the evidence when they formed their initial judgment. Does this lack of symmetry functionally distinguish it from the extra weight view?

It does not. Since the answer to the question of which party E tilts toward is itself dependent on who was right in their initial judgment, in attempting to ascertain and thereby properly respond to the tilt of E, both parties will have no alternative to re-deploying that same initial judgment. As Kelly puts it: There is no "warning bell" that goes off when you are mis-evaluating the evidence, and which lets you know that you're the one in the dispute who ought to lay down your arms; furthermore, the relevant facts may not seem to be facts at all, from your perspective.[22] So in trying to take account of the tilt of E, it's inevitable that one will de facto take it to tilt toward oneself. The total evidence view, just as much as the steadfast view, also functionally entails the extra weight view.[23]

The same is true of the right reasons view. On that view E does not merely tilt toward but is rather fully in line with whichever party was right in their initial judgment. So, for the same reasons, it will also functionally entail the extra weight view. The difference is just that it will functionally entail a particular weighting: namely, the extremal weighting that places all weight on one's own view and none on those of any interlocutors.

So, non-conciliatory views all functionally entail the extra weight view. When we forget about what people *should* do and focus just on what they *will* do, we see: people with non-conciliatory views will respond to the evidence E that arises in disagreement by acquiring

21. Enoch (2010, p. 989–990).

22. Kelly (2010, p.165, 167–172).

23. Indeed, this is why Setiya (2012) suggests, in friendly development of the total evidence view, that although the stronger status of justification should be reserved only for the correct party, the status of blamelessness should still be allowed to the incorrect. The incorrect party is trying to follow the correct epistemic norm — it's just that their situation is hopeless, and so blameless. As with Enoch, this again acknowledges the functional entailment while holding out that there is a significant non-functional epistemic difference; here the difference is between mere blamelessness and actual justification.

just the same attitudes that one would acquire by putting some extra weight on one's own view, taken as such.

Given that there is some, how *much* extra weight exactly is functionally entailed by these views? Misty starts out, before the final, taking things to be 50 / 50. How far away from that will she get, on the basis of E? Let's name that distance, whatever it is, N. Misty gets up to 50 + N / 50 - N.

In introducing 'N', I should flag what I do and do not assume about it. My assumptions are minimal. So: I make no assumption that there is some N constant across these views — as already noted, what's distinctive about the right reasons view is that it, unlike the other views discussed, seems to functionally entail a very high N. Nor do I assume that N will be constant across different developments of the individual views; you could for instance have different steadfast views that varied in how stalwart they commanded one to be. Nor do I assume that for each individual view it will assign some single N constant across different types and instances of disagreement: maybe N will depend in each instance on all sorts of variegated local facts. Nor do I assume that the actual value N will take for any given disagreement can be precisely deduced in advance.

There are, however, two things that I *do* assume. The first thing I assume is that N is sometimes non-trivial. Non-conciliatory views would not present a very interesting or exciting alternative if they posited some E such that it allowed Misty to be as much as an extra thousandth of a percent more confident in her correctness. Rather, the difference ought to be sometimes substantial. The second thing I'll assume is that, even if the value of N is not precisely knowable in advance of any particular disagreement, nonetheless a reflective and rational subject could at least estimate it.[24]

24. On some views, rational subjects will have perfectly sharp credences in even the most outlandish of propositions: see, for instance, Elga (2010b). That would be congenial to my argument, but I need not assume it. All I need to assume is that however well-behaved things need to be to yield an estimate that is itself well-behaved enough to figure in practical reasoning, N is at least that well-behaved. This is satisfiable both if N is poorly behaved but good behavior is not required, and if good behavior is required but N turns out to be well-behaved.

9. Extra Weight Decreases Expected Accuracy

So, the non-conciliatory views give characterizations of Misty's evidential situation on which they need not say that she changes her mind in response to the mere discovery that Ash disagrees. Rather, she changes her mind in response to E, evidence that she acquires along the way and which is described so as to look like the right sort of thing to rationalize her change of mind. Non-conciliatory views thereby distinguish their rational structure from the rational structure of the extra weight view. Nonetheless, they admit a functional entailment from their positions to the extra weight view. In this section I look harder at the consequences of that admission.

The question to ask now is: Given that E is as described, how should Misty see it as figuring into The Process? And the objection is: Given Misty's peerhood belief, and given the non-conciliatory views' functional entailment of the extra weight view, Misty should think that undergoing The Process with respect to E will reduce, rather than enhance, her accuracy.

We can illustrate this result with a flow chart. Suppose Misty is preparing to head off to her Frustrating Final. Misty is a non-conciliationist, and she knows it. She also knows that she and Ash are about to take a test with a single true-or-false question, and that they will compare their answers afterward. What are the relevant possible outcomes, and what does Misty now believe are the chances that each will occur?

There are two relevant branch points, and so three relevant outcomes. The first branch is that Misty and Ash may either get the same answer or they may get different answers. The second branch is that, provided they get different answers, Misty can be right and Ash wrong or Ash right and Misty wrong.

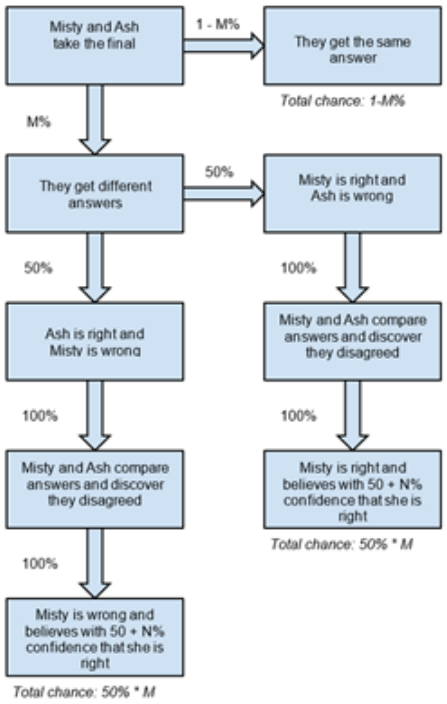
What are the chances on each horn of the first branch, as Misty now sees things? The chance that Misty and Ash will agree depends on

each of their independent reliabilities. Let 1 – M% be the chance that they agree, and hence M the chance that they disagree.

What are the chances on each horn of the second branch, as Misty now sees things? The second branch is over, should they disagree, which of the two is right. That one is straightforward: Misty takes that to be 50 / 50. That's just the content of her peerhood belief as we have defined it.

The final relevant feature of the diagram is that, on both horns, Misty winds up believing that she is 50 + N% likely to be right; such is the foreseeable functional entailment we outlined in the previous section.

Putting that together, we get the following:



Misty now believes that, if they disagree, it's 50 / 50 whether she or Ash will be right. But she also believes that if they do disagree, then in acquiring and responding to E she *will* become convinced that it's 50 + N% / 50 - N%, favoring her. Should she foresee that change producing an increase or a decrease in her expected accuracy? As I indicated earlier, my goal in this section is to argue that the answer is: She should foresee a decrease. I will defend that claim in two ways. First, with a lightly technical gloss, I'll note that it follows from the use of suitable accuracy measures. Then I will go on to offer some informal analogies which reinforce that verdict.

So, first: how should we try to quantify Misty's expectation? If she does in fact wind up being right, then the increase of N will wind up making her credence more accurate; if, on the other hand, she winds up being wrong, then the increase of N will wind up making her credence less accurate. Since she now thinks it's 50 / 50 which she'll get, whether this change has a positive expected value for Misty depends on whether the gain in accuracy that she gets in the good case is larger or smaller than the loss in accuracy that she undergoes in the bad case.

This will depend, in turn, on what "scoring rule" we use to measure accuracy. In general, we want to say that credences which are further from the truth are worse than credences which are closer to it. A scoring rule is a mathematical tool that tells us, given our credence and the actual truth, how to measure that distance between them, and thereby lets us draw the relevant comparisons.

Scoring rules can be divided into those which are "strictly proper" and those which are not. A scoring rule is strictly proper iff it renders all consistent credences self-endorsing with respect to expected accuracy. That is to say that, once we use some credence in setting the odds for an expected value calculation, the calculation then selects that very same credence as the one which uniquely maximizes expected accuracy. But notice that is just what we are asking about here: given that Misty takes it to be 50 / 50 who will be right, how accurate does she expect her later 50 + N / 50 - N credence to be? Adopting a strictly proper scoring rule settles the answer as: 'less, because given that we

are using 50 / 50 to set the odds for the expected value calculation, it will follow that 50 / 50 uniquely maximizes expected accuracy'. *Ipso facto*, 50 + N / 50 - N has a lower expected accuracy.[25]

So, given that our scoring rule is strictly proper, we can get the conclusion we set out to: Misty will see acquiring and responding to E as lowering her expected accuracy.

One might worry that by appealing to strictly proper scoring rules I rest my conclusions on an arbitrary choice in formalization. I have two responses: the first response is that the choice to use a proper scoring rule is not arbitrary, and in fact expresses no new commitment over and above those already introduced. In sections 1 and 2 I offered, as an anodyne starting point, the observation that we expect undergoing The Process to improve our accuracy, absent special reason to think otherwise, and that getting that increase in accuracy is part of the reason why we are concerned with acquiring new evidence. For this to be generally true, we already need to be committed to using strictly proper scoring rules.[26] The second response is that strictly proper scoring rules are useful to quantify my objection to non-conciliatory view, but that their only feature I've actually appealed to here — that random changes a fixed distance toward or away from the truth look bad epistemically — is anyway independently plausible.[27]

So, on reasonable ways of measuring accuracy, Misty should foresee a decrease in her expected accuracy. That is the first part of my argument. The second part reinforces that verdict by turning away from the abstract question of scores and directly considering *what Misty's epistemic situation is like*, on these non-conciliatory views. In answering, we can buttress our earlier, more abstract claims by giving some comparisons that make it clear both what is going on in Misty's case and why she should anticipate that undergoing The Process with respect to E is a bad idea.

25. So, for instance, take the Brier score, which is strictly proper. It measures accuracy indirectly by way of measuring inaccuracy (which we then minimize in order to become more accurate). To calculate the inaccuracy of a credence, it squares the distance between it and the actual value (0 if false, 1 if true). So, for a credence of .5, its inaccuracy will always be (1 - .5)^2 = (0 - .5)^2 = .25. A credence of .5 + N will have either inaccuracy (1 - (.5 + N))^2 or (0 - (.5 + N))^2, depending on whether the proposition is true or false, respectively. If it's taken to be 50 / 50 which, then the expected inaccuracy is .25 + N^2, and so this gain of N confidence would lead to an expected increase of N^2 inaccuracy.

26. To illustrate, consider a widely discussed improper scoring rule: the linear scoring rule. Under the linear scoring rule, the only credences which self-endorse from the perspective of accuracy are either maximally opinionated or perfectly indifferent. As it turns out, if someone follows the linear scoring rule, then any time they become less opinionated (without becoming perfectly indifferent), then by their lights both before and after the change, they expect their new accuracy to be lower than their old one. But surely sometimes it is rational to lose marginal confidence, as, for instance, when the new jobs report leads me to go from being .7 confident that the President will be re-elected to merely .6 confident. If the linear scoring rule is correct, there are going to be all sorts of rational belief changes such that the person undergoing them nonetheless anticipates their accuracy falling.

    The linear scoring rule presents a particularly dramatic example, but what goes for it also goes for improper scoring rules generally. Improper scoring rules force us to make a choice: either hold that considerations of accuracy are just normatively irrelevant, and so it is no problem that we sometimes expect ourselves to be doing poorly by our own accuracy-lights; or, alternately, allow that reflection on the nature of the satisfaction relation between probabilistic beliefs and the world actually substantially constrains the otherwise consistent beliefs we can reasonably hold and reason to — as, for instance, by ruling out all beliefs which are neither maximally opinionated nor maximally indifferent.

    The first option is ruled out by the opening remarks of the paper — I take it to be platitudinous that we take ourselves to ordinarily be improving our accuracy, and that doing poorly along our lights accuracy-wise is a problem. The second option is independently implausible. Reflection on the satisfaction relation may reveal that inconsistent beliefs are bad, but it is difficult to believe that it rules out whole swaths of consistent ones. C. f. Greaves and Wallace (2006, p.17–18).

27. So, for instance, Gibbard (2007) is skeptical of accuracy-first epistemology on the basis that the linear scoring rule seems to him to be a reasonable way of caring about truth, and it renders a hash of accuracy-first epistemology. Instead, he proposes that guidance value, rather than accuracy, is the appropriate candidate for the basic epistemic value. Although he thereby disagrees with the general methodological sympathies of the paper, his own view is nonetheless copacetic with the substance of the arguments, since his conception of guidance value also secures this feature.

So, what is going on? On each family of non-conciliatory view, Misty anticipates that undergoing The Process with respect to E will fail. But, as it turns out, it is for different reasons in each case.

For the steadfast view, Misty anticipates The Process failing due to misleading evidence. Whenever Misty anticipates a situation of peer disagreement, she thereby anticipates a situation where two pieces of evidence will be generated: one for her, one for the relevant Ash. These pieces of evidence will symmetrically support opposite conclusions about whose record has improved and whose has degraded. These Es by nature are generated in opposing pairs, and for each non-misleading E there is a paired misleading E.

Suppose that, instead of being ruled by a malevolent demon, we were ruled by an absurdist one. One day you run into the absurdist demon at a gallery opening and over free wine she lets it slip that she finds the idea of people coming to understand the world by way of reading newspapers to offend her sensibilities. And so what she does is this: every time some piece of non-misleading evidence is reported in a newspaper, she ensures that some other misleading newspaper report is also generated. The nature of this second report is that it always contains misleading evidence of the exact same strength as in the original report, and that this misleading evidence always indicates instead the opposite conclusion. Some newspaper reports evidence for P, which is true; she ensures that some other newspaper reports equivalently strong evidence for ~P, which, of course, is false.

If that happened, the first thing to think would be that opening the newspaper, reading the reports therein, and then believing them would lead to less accurate beliefs. It would be, for all you could tell ahead of time, random whether you were going to get one of the misleading or one of the non-misleading pieces of evidence, and random changes in one's beliefs are expected to degrade their accuracy.[28]

But the situation you would be in with respect to newspapers is just the same that we are all in with respect to disagreement, if the steadfast view is correct. When it comes to disagreement: for every non-misleading piece of evidence, there's an equal and opposite misleading piece, and antecedent reason to anticipate that you're as likely to get one as the other. And, unlike the newspaper case, there is no need for a demon to tip us off. It's an *a priori* feature of E, so described, that it must be this way.

For the total evidence and right reasons views, on the other hand, Misty anticipates The Process failing due to irrationality rather than misleading evidence. For these views, there is only a single E and it always points in the same direction as the correct assessment of the original evidence; there is no reason to think that it is systematically misleading. Yet there is nonetheless a reason to think that responding to it will systematically lower accuracy.

To see this, consider a new absurdist demon. This demon is such that her sensibilities are offended by your confidence in your rational faculties, and so she also has tampered with the newspapers. But this time, rather than mix in misleading evidence, she has instead ensured that all the reports are non-misleading. But she is still a villain, just more weirdly: when stocking the newspapers, for every piece of evidence that you will rationally appreciate, she has also inserted elsewhere a piece of evidence that you will irrationally misjudge. She has looked into your soul and seen all your prejudices and incompetencies, and she has thereby constructed a world full of good newspapers and good evidence — but yet nonetheless a world in which, for each time some newspaper reports evidence for P such that you consequently appreciate P, another reports evidence for Q such that consequently you mistakenly conclude ~Q.

This absurdist demon is different from the first, but her effect on your relationship to newspapers is the same. Again, if this were revealed, the first thing to think would be that opening the newspaper, reading the reports therein, and then drawing conclusions on that basis would lead to less accurate beliefs. It would be, for all you could tell ahead of time, random whether you were going to react rationally

28. As above, this will be true under any strictly proper scoring rule.

or irrationally to the evidence therein, and random changes in one's beliefs are, as before, expected to degrade their accuracy.[29]

And again, the situation you would be in with respect to newspapers in this scenario is just the same situation that we are all in with respect to disagreement, if the total evidence or right reasons views are correct. When it comes to disagreement: for every piece of evidence you'll rationally appreciate, there's a piece that you'll irrationally follow in an equal and opposite direction, and antecedent reason to think you're as likely to get one as the other. And again, no need for a demon to tip us off. It's an *a priori* feature of E, as so described, that it must be this way.

We opened this section with a chart demonstrating that since all non-conciliatory views functionally entail the extra weight view, all non-conciliatory views thereby construe disagreements as situations wherein attempting the proper response degrades expected accuracy. And now we've explained, for each family of view, why that's so. For steadfast views, it's because they construe the evidence acquired during disagreement, E, as such that it by nature is generated in equal and opposite pairs. The result, then, is that in responding to it one anticipates failures of The Process due to misleading evidence. For the total evidence and right reasons views, it's because they construe the evidence acquired during disagreement, E, as such that, by nature, for every rational response it provokes, it will also provoke an equal and opposite irrational one. The result, then, is that in responding to it one anticipates failures of The Process due to irrationality. For both, and for either reason, the relation to accuracy is negative.

10. Objecting to the Inference to Conciliation

Allow that all this is so: non-conciliatory views functionally entail the extra weight view, the reason for this is that the evidence non-conciliatory views appeal to predictably prejudices oneself toward one's own correctness, and the result is that one expects responding to that evidence to lower one's accuracy.

But what follows? My preferred answer applies the conclusions we drew in section 2: since we can anticipate that trying to respond to E-type evidence leads to a decrease in our expected accuracy, we ought not to respond to it. Instead, we ought to ignore it. And so, I think, we should be conciliationists. After all, once the E-type evidence is ignored, all that's left behind is the mere fact of disagreement, and all parties to the debate began by agreeing that *if* the mere fact of disagreement were the only epistemically relevant thing learned, then conciliation would be the only sensible response.

In this section, I want to get onto the table a way of resisting this conclusion. The response I'm going to consider accepts the earlier conclusions about the foreseeable consequences of trying to respond to E-type evidence, but disputes that we thereby ought to be conciliationists.

So far, I have treated *steadfast*, *total evidence*, and *right reasons views* in parallel, because the functional equivalence to extra weight holds of all. But the considerations I'm about to advance are no longer indifferent among them. Rather, I am going to put aside steadfast views, as the response I am going to consider is available only to total evidence and right reasons views. This response exploits the fact that these views conceive of E in such a way that it is biased toward the truth, and, as such, the functional equivalence with extra weight holds only because subjects expect that they will sometimes behave irrationally. Both total evidence and right reasons views make use of such a bias toward the truth, and because of this I will refer to them, from here on out, as BT views.[30]

The response has two parts. The first goes like this: It's true that you can anticipate that should you respond to some piece of evidence E, the credences you will form in so doing will be losers from the perspective of accuracy. But that is only because you anticipate that you

29. Again, this follows from using a strictly proper scoring rule.

30. I borrow the phrase 'bias toward the truth' from Setiya's development of his (2012) non-conciliatory view.

may make a rational error, and, in general, the badness of the errors you make is going to outweigh the goodness of the successes. But in those cases where you *don't* make a rational error, there is no argument that responding to E is bad; rather, the platitudes about the general goodness of responding to evidence should lead us to expect the opposite. Responding to E without making a rational error improves accuracy. And, as a proponent of a BT view, I have always thought that what you *should* do is respond to E *without* making any rational errors. No doubt has been cast on the goodness of *this* policy.

The second part continues: Furthermore, not only is *successfully* following my norm (the only type of following I recommend) a good deal with respect to accuracy, but someone who is doing so may recognize themselves as so doing, and thereby need not worry about the general tendency of E-type evidence to lead to error. We may grant that Misty should expect that, at the level of description 'some item of E-type evidence', responding to that item of evidence will lead her astray. But in A Frustrating Final she doesn't receive "some item of E-type evidence". She receives *this* item of E-type evidence. And she may correctly perceive that *this* item of evidence supports her. If she does so, she may maintain her belief that *in general* items of type E will lead her astray, but still take *this particular* item of evidence to support her correctness. So one can acknowledge the general claims about E-type evidence thus far argued while still following a BT view and fully endorsing the results they get by doing so. This is because, in each particular case, successfully following the rule will involve genuinely recognizing that they are getting it right.[31]

Now, of course, sometimes the people trying to follow these rules will make a mistake and instead go wrong, and worse yet, when they do so they will also mistakenly think that they are getting it right. But this regrettable behavior is no mark against those who are instead epistemically excellent.

31. Thanks to an anonymous referee for pressing me on the distinction between conclusions about E-type evidence generally and conclusions about any specific piece of E-type evidence.

My goal over the next sections will be to dispel the appeal of this line of response. One strategy for doing so focuses on the last concession: we might think that theories of epistemic rationality shouldn't be allowed to so cavalierly wash their hands of those who try, but fail, to live up to their standards. That will not be my focus. Rather, I want to attack BT views on their most favorable grounds. So I will, in pursuing my response, focus exclusively on the best case for the BT views: the case where Misty really has evaluated every piece of E-type evidence she's received rationally, and where she in each case has correctly ascertained that she has done so. Even in this case of perfect performance, I think the BT view significantly underestimates the difficulties of combining a recognition of one's general fallibility in handling E-type evidence with a policy of uncritically responding to it in each particular instance. To draw out these difficulties, I start by considering what life would be like for someone who accepted my general conclusions about E-type evidence, but then tried to hold on to their non-conciliatory view in this way. My first claim will be that such a person is committed to pervasive self-binding; illustrating that will be the task of the next section.

11. BT Views Lead to Self-Binding

My general strategy for illustrating the self-binding that BT views require will be to return to our central example, of Misty taking A Frustrating Final, and to consider a series of bets that we might offer her on whether she or Ash has gotten the right of things. As we have established, earlier in The Process she anticipates that responding to the E-type evidence she is bound to receive will degrade her accuracy. It follows that she thereby expects decisions she makes on the basis of those later credences to do worse. Since she expects decisions based on her later credences to do worse, she should now be willing to do her best to preempt them. So, for instance, she should be willing to pay money in order to constrain or eliminate choices she might later be offered. As such, she will be interested in binding her future self to her current will.

To illustrate: suppose Misty has thought things through and come to adopt the BT-based response I have outlined. She accepts that, in general, trying to respond to E-type evidence will lower her accuracy, but nonetheless holds a BT-type view and thinks that, in each instance, what she ought to do is respond to E *correctly*. Now also suppose that I am lurking about, waiting to make trouble. I tell Misty, "I am going to watch you and Ash leave the lecture hall after your final examination. I will wait for you to compare your answers, and then — if I see the look of mutual dismay on your faces that indicates you disagreed — I will spring from the bushes! And I will offer to you a deal. I will sell you a ticket that pays out a dollar if, once the grades come in, it turns out you were the one who was wrong on that final. I will offer it to you at the low, low price of 50 - N/2 cents."[32]

Misty knows she will reject that offer. She knows that in the situation described, if it occurs, she will think that she is 50 + N% likely to be right. Therefore, at the time that she is offered this bet she will think the expected value of the ticket is only 50 - N cents, and so decline the opportunity to purchase it.

But that is then and this is now. Currently, Misty still thinks it's 50 / 50 whether she'll be the wrong one should a disagreement arise, and so thinks that the ticket has an expected value of 50 cents. As such, she thinks that although she unfortunately *won't* be willing to buy it, if she *were* to do so she would net an expected gain of N/2 cents. Misty thinks that if I do what I say, the result will be that she misses a good deal.

But if that's really what she thinks, it makes sense for her to act to try to secure that good deal. For instance, Misty could counter-propose to me the following, slightly more complicated bet: First, she pays me any amount up to M% of N/2 cents. Then, if she and Ash wind up disagreeing, she must pay me an additional 50 - N/2 cents. Thereafter, when the grades come in, if it turns out that Misty was the one that was wrong, I pay her a dollar. Why should she be willing to make this counter-proposal? Because on her current understanding of the situation, it nets an expected profit of whatever difference remains between her initial payment and M% of N/2 cents.

That's all a mouthful. But the informal description of what Misty is doing by way of this counter-proposal is straightforward. Misty is paying me some money now in exchange for my later forcing her to take that bet, the bet she otherwise would decline. Since the amount of money she's paying me is smaller than the expected profit from that bet, she should think that, even minus the payment, she'll still net a gain. Current Misty is not willing to leave things up to future Misty; she wants to lock in her choice now, because she knows what future Misty *will* believe, and she currently takes it to be a worse basis for decision-making than her current credences.

12. Self-Binding and Conciliation

If Misty accepts my characterization of E-type evidence, yet nonetheless holds on to her BT view, she will thereby be committed to self-binding. This is not yet a *reductio* of her position. Those with a fondness for BT views may take the argument thus far not to refute them, but merely to be an interesting discovery: it turns out, because BT views are true, that cases like A Frustrating Final are cases where self-binding is rationally required. Call this the BTS view, as it is the BT view as supplemented by self-binding.

We can illustrate this thought with an example from Roger White, who himself appears to hold a BTS-style view;[33] It may make sense at the beginning of the night, given how much you anticipate drinking, to give your keys to your friend and tell them not to let you drive home. You may do this because you think that later in the night you will be drunk, and there's a significant risk that you may misevaluate your state and consequently try to drive. But suppose, contra your expecta-

32. One might object: Money is not continuously graded, so if N is small enough there's no guarantee that there will be appropriate cent values to make the argument go through. It would be curious, however, to argue that non-conciliatory views are defensible because the money we contingently use is coarsely graded.

33. White (2011, p.601–604).

tion, the night passes tamely and at the end of it you are rather sober. You may not only *be* sober, but *know* that you are. If so, then you may rationally seek to do what you earlier rationally tried to restrict yourself from doing, i. e., get your keys from your friend and drive home.

The BTS view assimilates cases of peer disagreement to this model. It holds: *in general* E-type evidence is such that responding to it reduces expected accuracy. And for that reason it makes sense, ahead of time, to take measures to prevent yourself from making decisions on its basis. But once you get any particular piece of E-type evidence, you may not only respond rationally to it, but you may furthermore see that you have done so. And once you see that, you may rationally go on to try to make the very decisions you earlier were rationally trying to restrict yourself from. There is a sort of practical friction here, but the epistemic asymmetry underlying the practical friction is well-motivated, and so, the thought goes, there's nothing wrong with it.

Now, before going on to give my argument against BTS views, I want to remark on why, even if they are in the end correct, they still preserve a significant conciliatory spirit. Much of the interest that drives the peer disagreement debate is practical. As such, if it turned out that we ought to make sure we conciliate — not because it is rational in the moment, because it isn't, but rather because it is rational to make ourselves into the sort of creatures who will be constrained to conciliate in the moment — it would still be true that we should try our very best to see to it that we go on to conciliate. And learning that we ought to make ourselves into conciliators would have deep practical consequences.[34]

So I think it is important to recognize that BTS-licensed conciliatory self-binding, when genuinely pursued, may result in a surprisingly robust conciliatory program. In pointing this out, though, I do not mean to suggest that BTS views will wind up having identical practical upshots to conciliatory views: exactly how close they wind up depends on the answers to questions I won't enter into here.

I should note, however, one thing that I *do* assume about the self-binding that BTS views endorse. Namely, I assume that it is not adequate to bind oneself merely by making a conscious decision to set a policy. I take it that this is what gives the BTS view its distinctive non-conciliatory flavor. If all it took to bind ourselves to a policy was to select it by some conscious act, then the difference between a program BTS-licensed conciliatory self-binding and conciliation proper would risk becoming slim indeed. In any case, I am going to assume that the sort of binding a BTS-er imagines as effective is e. g. the sort relayed in White's story where one commissions a friend to act as controlling supervisor of one's keys, or in my earlier discussion where one uses pre-betting to effectively take decisions out of one's future hands.

13. BTS Views Deprive Us of the Benefits of Our Faculties

Why not hold a BTS view? In this section I argue that BTS views construe our epistemic predicament as tragic. I hold, however, that we have no compelling reason to accept this tragedy. So we ought to reject it, and BTS views along with it.

What is the alleged tragedy? If BTS views are true, then there are cases in which Misty thinks her powers of judgment are a valuable indicator of the truth about a question she is deeply interested in, yet nonetheless she will rationally do her best to avoid deploying them. She will choose instead to carry on in ignorance. Rationality will thereby require her to lock herself out of the benefits of her own faculties.

The case I'll use to demonstrate is similar to the earlier Frustrating Final insofar as it involves Misty thinking ahead of time about a future disagreement; it is more complicated, however, insofar as we consider a situation in which more than one peer is at work. So consider the following case:

34. Compare with Newcomb cases: if we think two-boxing is rational but nonetheless we prudentially ought to see to it that we become one-boxers, this conclusion is not very practically compelling. After all, we do not ever expect to face a Newcomb case.

A Crowd of Experts: Misty holds the BTS view. She is also an expert mathematician, and she and her eight friends work in an obscure subfield. Recently, she has become aware that they are abuzz over a new result; four of them think that the proof of this result is valid, and four of them are convinced that it relies on a subtle equivocation. Misty doesn't yet know what the result is or how the putative proof goes, but she thinks of herself and her friends as independent and highly reliable judges; given that they are evenly divided in this early stage of investigation, she takes it to be 50 / 50 whether the proof is valid. Misty, anticipating that she will soon be asked to investigate the proof herself, considers what will happen when she does. She anticipates that she will either find it valid or find it invalid. Conceiving of this as a future exercise of her mathematical judgment, in the abstract, she thinks of it as on a par with the judgments of any of the other experts who have already evaluated the proof. So, she anticipates that there will be a 5–4 split of the relevant experts, and she thinks that, under those conditions, the odds rise to 65 / 35 in favor of whichever option she judges, given that it will ipso facto command the slight majority. She also thinks that she will, after forming her judgment, not merely respond to the fact that she so judged, but also respond to the richer E-type evidence that will then become available to her. She anticipates that she will, in so doing, become predictably convinced that it is 85 / 15 in favor of the option she judges (that is to say, the predictable extra weight she will accrue in the course of applying her BTS view will amount to another 20% confidence — or, in my earlier terms, N is .2).[35]

35. The situation so described is not intended to be a general description of mathematical epistemology, e.g. there is no reason to suppose that it will always be true that one ought to suppose, when mathematicians evenly divide over the validity of an early proof, that it is thereby 50 / 50 whether it is valid, nor is there any special connection between a 5–4 split and the particular 65 / 35 or 85 / 15 odds assigned. The numbers in this story are not supposed to be deduced from one another; they are just values which we have no reason to think couldn't arise.

Now consider the following development, wherein we put a practical price on inaccuracy:

Malevolent Aliens Force Practical Consequences: Misty is abducted by powerful, malevolent, and mathematically sophisticated aliens. They inform her that they have long ago considered the mathematical result being discussed by her colleagues, along with the attendant line of proof, and they know with perfect certainty whether it holds. They offer to show her the purported proof her colleagues are puzzling over and give her a chance to think it through. They also inform her that, in a week, they will ask her how likely she thinks it is that the proof is valid. They will then take her answer, measure its inaccuracy using the Brier score, and multiply the result by 10,000 and murder that many humans. She has every reason to believe this is true.

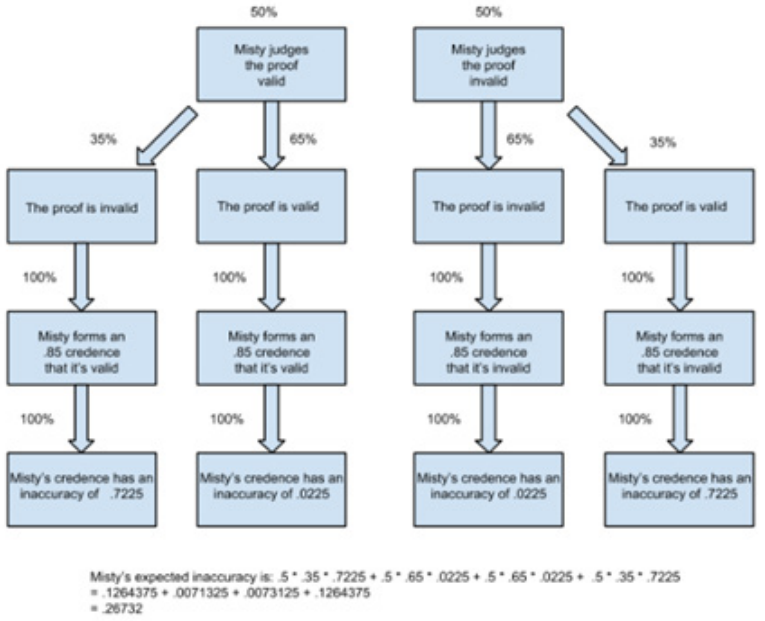
Finally, suppose:

No Restraints: Stranded as she is, Misty has no available means to bind her future self to any current decisions. There's no time to instill habits; the aliens are not susceptible to elaborate insurance bets; she cannot pre-select answers, etc., etc. Her only choice is over whether or not she wants to see the purported proof before she reports on whether it is sound.

In the case as described, Misty ought to decline to see the purported proof. Given that she holds a BTS view, she anticipates that she will predictably accrue extra weight to her assessment. And in this case, that extra weight is such that her expected accuracy afterward will be

worse than her current 50 / 50 guess. As such, in declining to see the proof, she minimizes the expected death count.

We can illustrate this with another (hideous) flowchart, again depicting what Misty anticipates will happen if she is shown the purported proof. She currently takes it to be 50 / 50 whether the proof is valid. Given that she is currently indifferent about the validity of the proof, something which she takes her judgment to be an imperfect indicator for, she is also indifferent about whether she will judge it valid or invalid. So the first branch in our chart represents the 50 / 50 chance she assigns to her judging the proof valid. From there, we add the next branch, representing how like she takes it to be that the proof *is* (in) valid, given her judgments; the odds of these branches are 65 / 35 in each case that she is getting it right, as that's what she takes the 5–4 split of experts that will then exist to indicate. And then we can drop down a final row of boxes that shows that, in all cases, she expects to form a .85 credence that whichever option she has judged is correct; this is the additional extra weight that she will accrue to her view in attempting to respond to the richer evidence available to her when she sees the actual proof. Finally, that .85 credence is labelled with how inaccurate it is by way of Brier score in each case:



Multiplying out the branches of the flowchart, we see that Misty's expected inaccuracy after being shown the proof is .26732. Misty's current inaccuracy, given her 50 / 50 credence, is .25. So Misty's expectation of her inaccuracy upon seeing the proof is .1732 worse, corresponding to a worsening of the expected death toll by 1,732 people.

I have constructed this example with particular numbers, and I have used a particular scoring rule — the Brier score — but the structure is general. If BTS views are true, we should expect such cases to arise. What generates the accuracy deficit is the predictable extra weight Misty expects to accrue, and that there is such predictable extra weight is just the same core feature of non-conciliatory views we already explored in the context of A Frustrating Final.

Rather, what is interesting and new about the way we've filled in the case is that we've set the value of N — the predictable extra weight Misty will accrue — relatively high. The extra weight here is pretty

heavy.[36] And that is the important difference with A Frustrating Final. In A Frustrating Final, we showed that the evidence E non-conciliatory theorists appeal to is such that responding to it lowers Misty's expected accuracy, and hence, if she could choose ahead of time to self-bind to avoid responding to it, she would. However, in that case the E-type evidence she receives comes packaged with some other evidence — namely, the facts about distribution of expert judgment that she acquires over the course of thinking through the problem and then meeting up with Ash. So, for all we have said, it may well be that even though the E-type evidence alone has a negative effect on her expected accuracy, that's outweighed in the final balance by the rest.

By contrast, here we have fixed the pernicious effects of the E-type evidence to be significant enough that Misty not only anticipates that responding to it will lower her expected accuracy, but that this bad effect will be large enough to outweigh the other benefits Misty gets from working through the problem and adding her expert judgment to the stock of those already existing. So Misty should not want to work through the problem and add her expert judgment to the stock of those already existing: she should prefer to avoid thinking about it altogether, and just answer the aliens on the basis of what she already knows about how the *other* experts have answered.

On reflection, however, we should see this as a perverse result. After all, Misty is a mathematical expert. Everyone, herself included, takes her judgment of the purported proof to be a genuine indicator with respect to its validity. And yet, when it comes time to make this important decision, in which many lives depend on the validity of the purported proof, she'd rather not draw on her own judgment. She'd rather make her decision *only* on the basis of the eight other expert mathematicians' judgments.

She does not think those other experts' judgments are any better than hers, and when she thinks about expert judgment in the abstract, more judgments are always better. If she could have chosen for another expert to have seen the proof, judged it, and then had their opinion be reported to her, then she certainly would have. But despite being just such an expert herself, she cannot make use of her own judgment.

And this, I take it, is the sense in which she is deprived of the use of her own faculties. Because she subscribes to the BTS view, Misty must leave her powerful and relevant mathematical abilities fallow, even as she is confronted with precisely the sort of problem to which they are so well-suited to contribute.

14. BTS Views' Tragedy Is Undermotivated

As an acolyte of the BTS view, Misty will decline to see the proof. But what if she winds up seeing it anyway? Suppose the aliens show it to her against her wishes. If so, then she will evaluate it and come to be .85 confident in its (in)validity — as laid out in the previous diagram. In this section I want to suggest that there is something very odd about this combination.

Return to an earlier example: I have left some possessions stranded at the house of my ex-boyfriend. I want to get them, but I also know that he is an expert at manipulating me and so do not want to give him the chance to speak to me.

How should I think about his hypothetical speech in advance? I should think that I will almost certainly not respond to it rationally. Even though I am highly confident that nothing he says will present a compelling reason to believe him, I still think there's a significant chance I will believe him anyway. So I anticipate that my response to hearing him out will be irrational in a disastrous way.

At the same time, I should also think that his speech will consist in perfectly good evidence. After all, the content of his speech will likely give a picture of his state of mind and show with greater clarity the

36. Perhaps this is illicit? I think not. In cases of deductive reasoning, like the one we are presently considering, numbers like these make sense; deductive reasoning will often be such that one's individual judgment considered as such is a weak indicator, but the first-order considerations are very powerful — deductive arguments are, after all, as strong as first-order considerations get. So the difference between the conciliatory response and the response that is justified by the first-order considerations should be very large, and so N should be able to range quite high.

emotional tactics he uses to try to control me. And, though it is long odds, there is even the remote chance that his speech would contain genuine evidence that I ought to do and believe as he wants — it is at least *conceivable* that there could be such a reason, and that he could explain it to me.

Given this set-up, ask: Should I want my beliefs to be informed by responding to my ex's speech? There is a tension here. On the one hand, it's true that the speech is in some sense good evidence. On the other, though, I anticipate that my response to it will nonetheless be very bad. Here are two coherent ways to resolve this tension:

First, we might emphasize that the speech is good evidence. As such, we may say: Of course I should want my beliefs to be informed by responding to it. So I should make sure that I get the chance to hear it; I should seek him out and listen attentively. Once I've listened to his speech, I should then carefully think it over and apportion my beliefs as closely to this new evidence as I possibly can. I should always want my beliefs to be informed by good evidence. I find this answer very difficult to believe in light of the case — as I've already made clear in section 2 — but nonetheless it at least presents a consistent view.

Second, we might emphasize that I anticipate my response being bad; I am very confident that I will be convinced to believe something terrible. As such, we might say: I ought not to want my beliefs to be informed by responding to this evidence, because that will likely make them terrible. So I should make sure that I do *not* get a chance to hear my ex's speech. And supposing that somehow, against my will, I do manage to hear it, I should make sure that I don't let it influence me. If he manages to accost me, I should ignore him. I should ignore my own impulse to believe him, and so on — at least best as I can. I think this, as I have again made clear, is the better answer.

But regardless of which is better, both answers are consistent in the following way: they give a unified answer to *whether I should want to hear the speech* and *whether I should listen to it if I do*. They answer, naturally, in terms of whether I should want my beliefs to be informed by my ex's speech. If yes, then, as in the first response, I should both seek out and listen to it. If no, then, as in the second response, I should both avoid and do my best to ignore it.

The BTS view, by contrast, splits its answers. It tells Misty: Make sure you don't see the proof if you can help it. But once you do, you should carefully evaluate it and apportion your beliefs as appropriate. But this mix looks odd when we hold it up against the purer alternatives. Who would say: You ought to avoid your ex, surely, but if he manages to catch and corner you, then you should carefully evaluate what he has to say?

After all, it seems that the very same reasons I have for not wanting to run into my ex count just as much in favor of trying to ignore him once I do. And given that the very same reasons are operative for each — reasons relating to our anticipated rational failures — we can even think, if we like, of our ability to ignore as an internal counterpart of external avoidance strategies. One way we can deal with the fallibility of our rational capacities is by avoiding the sort of evidence that we anticipate will lead us astray altogether, as, for instance, when we try to avoid coming into contact with manipulative exes. But another way is by allowing ourselves to encounter the evidence, but then exercising higher-level control to stymie our responses to it, as we do when we encounter those exes but do our best to muffle our impulses to believe them. The BTS view fails to notice the continuities in both purpose and effect of these strategies.

The crowd-of-experts case is designed to highlight the arbitrary and inadequate nature of purely external management. In that case, Misty's external management strategies are limited to two options: either she can see the purported proof, or she cannot. If she sees the purported proof, she will get both the evidence consisting in her expert judgment of its validity, and also the richer E which includes whatever counts as the first-order evidence in a case of mathematical proof. She can't get one without the other, because she can't generate her expert judgment of the purported proof's validity without actually wading through the relevant E. So she has no way to use external management to achieve the result that she anticipates would actually be best:

getting an additional expert judgment on the soundness of the proof without accruing all the extra weight from that E.

If she allows herself to practice internal management, though, she can. She can get the benefit of her expert judgment without being swayed by the co-occurring E-type evidence: she can do so just by using that evidence to form her judgment, but then ignoring it thereafter.

As far as I can tell, BTS views have precisely one objection to this internal management: when Misty ignores the full force of her evidence, she misses out on the benefits she could get from it. But it is worth remembering that, when we look at Misty's passage through the whole of the case, *regardless* of which view she follows she will lose out on the full potential benefits of incorporating E into her beliefs. After all, on BTS views, when things go according to plan, she will never even get E in the first place. Rather, she will decline the aliens' offer to see the proof. BTS views need it to be the case that there is a distinction between two *ways* of excluding E from being factored into her final beliefs, and they need this distinction to be significant enough to justify pursuing purely external management even when internal management would allow for better results. This is hard to sustain when, theoretically, external and internal management share common purposes and effects, and intuitively, when we look to cases like that of the manipulative ex, the strategies seem to go hand in hand.

Thus concludes my brief against BTS views. On such views, Misty's situation is epistemically tragic insofar as she cannot make good use of her own expertise. This tragedy rests on a distinction between internal and external tactics for managing one's beliefs. I say we ought to reject both the distinction and the tragedy. Misty ought to get use out of her faculties, and she ought to do so by managing their deployment in the way the conciliationist suggests.

I spent this time going over BTS views because they represent, I think, the best chance for conceding that E-type evidence is deleterious to expected accuracy, yet blocking the further inference to conciliation. Having argued that BTS views fail, I thereby take myself to be entitled to reinstate that inference. E-type evidence is deleterious to expected accuracy, and so we ought to ignore it. Once we do, all that's left to respond to is the mere fact of disagreement — and once we limit ourselves to that, we are conciliationists.

15. Reflection and Conditionalization

We have traversed a great deal of ground. It is worth taking a moment to situate my argument in relation to some other, perhaps more familiar, epistemic claims and arguments.

I have argued that non-conciliatory views take what they foresee to be bad epistemic deals, and that these deals look bad as a result of a disagreement between Misty's epistemic point of view at two different times. Right now Misty thinks there's a 50 / 50 chance that if she and Ash disagree tomorrow she will be right, but at the same time she thinks that if she and Ash disagree tomorrow she *will, at that time,* believe that there's a more than 50 / 50 chance she's right. Since she does not now accept those odds that she knows she will later accept, she perforce thinks the later odds are less accurate. This generates the funny behavior.

One might think, then, that what's needed is a principle guaranteeing regularity between present and future beliefs. A commonly discussed family of such principles are *reflection* principles.[37] For instance, one such principle might be: If you have good reason to think that tomorrow you will believe *p,* believe *p* now. After all, suppose I tell you, "Tomorrow you will think it's raining." Then isn't that a good reason to think it will rain tomorrow? After all, usually you think things because they're so. So, we might ask, can the force of the argument given in this paper be captured by such a reflection principle?

The answer is no. All plausible versions of reflection principles include exemptions for, among other things, irrationality. Suppose I tell you, "Tomorrow you will be captured and brainwashed into thinking Obama is a lizard person." This is not a good reason to now think

37. For an argument against steadfast views that does go by way of reflection, see Setiya (2012). For an original locus of discussion on reflection principles, see van Fraassen (1995).

Obama is a lizard person.[38] As such, reflection principles will demand matching between present and future only when one has good reason to think that one's future beliefs will be rational. But, at least on the total evidence and right reasons views, this will not be so in cases of peer disagreement: one will think, rather, that it's only 50 / 50 whether one's future attitude will be rational. So even if some reflection principle is true, it will not be able to tell against those views. Yet the argument in this paper does tell against those views. Therefore, the argument in this paper cannot be captured by some form of reflection principle.

I have put my critique in terms of expected accuracy. Such a framing might lead one to wonder: There already exist arguments in the literature which purport to show, with mathematical clarity, that expected accuracy is uniquely maximized by holding a probabilistically coherent set of credences and then updating them by conditionalization.[39] Together, these two claims — probabilism and conditionalization — constitute the traditional Bayesian package, and so, it might be thought, these arguments show that anyone concerned with maximizing their expected accuracy ought to become some stripe of Bayesian. But if this is so, I am in trouble, as investigation reveals that traditional Bayesianism is actually inconsistent with conciliation.[40] If Bayesians have a mathematically sound monopoly on expected accuracy, then in trying to use expected accuracy to argue for conciliation I must have gone terribly wrong somewhere.

The answer is that those arguments do not actually show that anyone concerned with maximizing expected accuracy ought to become some stripe of Bayesian. Take conditionalization: such arguments may show that *actually* updating by conditionalization maximizes expected accuracy. However, they do not show that *trying* to update by conditionalization maximizes expected accuracy. And in many cases we have decisive evidence that if we *try* to update by conditionalization, what we will in fact do will be something else entirely.

We can again frame this in terms of the examples from the beginning of this paper. Suppose I listen to the testimony of my ex-boyfriend, or view the races and genders of all the applicants auditioning for my orchestra: I have strong evidence that if I do so, I will not react by conditionalizing on that new evidence. Rather, I have strong evidence that I will react by being convinced to return to my ex, or by forming more negative evaluations of the minority applicants' talents — and those are beliefs I now anticipate to be less accurate than my current ones.

I take it that the fact that *succeeding* at conditionalization maximizes expected accuracy is of great epistemic interest. Still, there are lots of things it would be lovely to succeed at, but that it's nonetheless best not to try — because one will likely fail, and the costs of failure will be significant. For instance, even if actually doing a backflip would impress everyone in the room, I ought not try. I take it that when considering what we should believe, just as when considering what we should do, we ought to take account of evidence that we will not succeed at doing what we try. In these cases, I have excellent evidence both that I will not succeed at conditionalization and that the results will be bad. So I should avoid trying to conditionalize on that evidence.[41]

16. Conclusion

When Misty finds that Ash disagrees with her over the answer to their Frustrating Final, what should she believe? Should she take things to be 50 / 50, as she antecedently expected they would be in light of his disagreement, or does she need to adjust that expectation in light of further evidential features of the case? Conciliationists and their

38. See Christensen (1991) for a compelling and detailed presentation of this point.

39. For the argument for probabilistic consistency, see Joyce (1998); for the argument for conditionalization, see Greaves and Wallace (2006).

40. C. f. White (2009) on the incompatibility of the "calibration rule" with Bayesian epistemology.

41. Here I am in substantial agreement with Schoenfield (2015). I take it that the rules it would be best to follow and the rules that it would be best to try to follow will each play substantial roles in our total epistemic theory — and, furthermore, that it is an attention to the latter which motivates conciliation. Thus I agree that conciliation ought to be grounded in a 'trying' account; I take it such an account will govern the prescriptive 'should' of 'should believe'.

opponents have clashed over the proper characterization of the evidence at Misty's disposal, under the assumption that answering the question of what her evidence supports would straightforwardly answer the question of what she ought to believe. If her evidence is exhausted by the mere fact of disagreement, then conciliation stands; if her evidence outstrips the mere fact of disagreement, then conciliation falls.

Evidence is being afforded a central role in the debate, but it is worth taking a step back and asking why it is that we care about evidence in the first place. I have framed the positive role of evidence in terms of a process of inquiry we value for its tendency to lead us toward truth and away from error. But if this is really what is valuable about getting and responding to evidence, then conciliatory answers to what Misty ought to believe can survive even quite non-conciliatory construals of the evidence available in disagreement cases. I have argued that construing such evidence as having anti-conciliatory force at the same time makes the evidence such that Misty expects trying to respond to it to lead her into error. So, from her perspective, any such anti-conciliatory evidence thereby lacks the truth-conduciveness that makes evidence worth paying attention to in the first place.

And that, then, is my ultimate conclusion. If there is any anti-conciliatory evidence, then it is highly unusual in precisely such respects that we should not want to respond to it. So we shouldn't respond to it. We should conciliate.

Works Cited:

Christensen, David. "Clever Bookies and Coherent Beliefs." *The Philosophical Review* 100, no. 2 (April 1991): 229–247. https://doi.org/10.2307/2185301.

Christensen, David. "Epistemology of Disagreement: The Good News." *The Philosophical Review* 116, no. 2 (April 1, 2007): 187–217. https://doi.org/10.1215/00318108-2006-035.

Christensen, David. "Disagreement as Evidence: The Epistemology of Controversy." *Philosophy Compass* 4, no. 5 (September 1, 2009): 756–767. https://doi.org/10.1111/j.1747-9991.2009.00237.x.

Christensen, David. "Disagreement, Question-Begging, and Epistemic Self-Criticism." *Philosopher's Imprint* 11, no. 6 (March 3, 2011). http://hdl.handle.net/2027/spo.3521354.0011.006.

Elga, Adam. "Reflection and Disagreement." *Noûs* 41, no. 3 (September 1, 2007): 478–502. https://doi.org/10.1111/j.1468-0068.2007.00656.x.

Elga, Adam (2010a). How to Disagree About How to Disagree. In Ted A. Warfield & Richard Feldman (eds.), *Disagreement*. Oxford University Press. 175–186.

Elga, Adam. "Subjective Probabilities Should Be Sharp." *Philosopher's Imprint* 10, no. 05 (May 20, 2010b). http://hdl.handle.net/2027/spo.3521354.0010.005.

Enoch, David. "Not Just a Truthometer: Taking Oneself Seriously (but Not Too Seriously) in Cases of Peer Disagreement." *Mind* 119, no. 476 (October 1, 2010): 953–997. https://doi.org/10.1093/mind/fzq070.

Gibbard, Allan (2007). Rational Credence and the Value of Truth. In Tamar Szabó Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology: Volume 2*. Oxford University Press. 143–164.

Goldin, Claudia, and Cecilia Rouse. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90, no. 4 (September 2000): 715–741. https://doi.org/10.1257/aer.90.4.715.

Greaves, Hilary, and David Wallace. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind* 115, no. 459 (July 1, 2006): 607–632. https://doi.org/10.1093/mind/fzl607.

Joyce, James M. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65, no. 4 (December 1, 1998): 575–603. https://doi.org/10.1086/392661.

Kelly, Thomas (2005). The Epistemic Significance of Disagreement. In Tamar Szabó Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology: Volume 1*. Oxford University Press. 167–196.

Kelly, Thomas (2010). Peer Disagreement and Higher Order Evidence. In Alvin I. Goldman & Dennis Whitcomb (eds.), *Social Epistemology: Essential Readings*. Oxford University Press. 183–217.

Kelly, Thomas (2013). Disagreement and the Burdens of Judgment. In David Phiroze Christensen & Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press. 31–53.

Lackey, Jennifer (2010a). A Justificationist View of Disagreement's Epistemic Significance. In Alan Millar, Adrian Haddock, & Duncan Pritchard (eds.), *Social Epistemology*. Oxford University Press. 298–325.

Lackey, Jennifer (2010b). What Should We Do When We Disagree? In Tamar Szabó Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology: Volume 3*. Oxford University Press. 274–293.

Lam, Barry. "On the Rationality of Belief-Invariance in Light of Peer Disagreement." *The Philosophical Review* 120, no. 2 (April 1, 2011): 207–245. https://doi.org/10.1215/00318108-2010-028.

Lam, Barry. "Calibrated Probabilities and the Epistemology of Disagreement." *Synthese* 190, no. 6 (April 1, 2013): 1079–1098. https://doi.org/10.1007/s11229-011-9881-0.

Schafer, Karl. "How Common Is Peer Disagreement? On Self-Trust and Rational Symmetry." *Philosophy and Phenomenological Research* 91, no. 1 (July 1, 2015): 25–46. https://doi.org/10.1111/phpr.12169.

Schoenfield, Miriam. "Chilling Out on Epistemic Rationality." *Philosophical Studies* 158, no. 2 (March 1, 2012): 197–219. https://doi.org/10.1007/s11098-012-9886-7.

Schoenfield, Miriam. "Bridging Rationality and Accuracy." *The Journal of Philosophy* 112, no. 12 (December 1, 2015): 633–657. https://doi.org/10.5840/jphil20151121242.

Setiya, Kieran (2012). *Knowing Right From Wrong*. Oxford University Press.

Titelbaum, Michael (2015). Rationality's Fixed Point (Or: In Defense of Right Reason). In Tamar Szabo Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology*. Oxford University Press. 253–294.

van Fraassen, Bas C. "Belief and the Problem of Ulysses and the Sirens." *Philosophical Studies* 77, no. 1 (January 1, 1995): 7–37. https://doi.org/10.1007/BF00996309.

van Wietmarschen, Han. "Peer Disagreement, Evidence, and Well-Groundedness." *The Philosophical Review* 122, no. 3 (July 1, 2013): 395–425. https://doi.org/10.1215/00318108-2087654.

Weatherson, Brian (2013). Disagreements, Philosophical and Otherwise. In David Christensen & Jennifer Lackey (eds.), T*he Epistemology of Disagreement: New Essays*. Oxford University Press. 54–73.

Weatherson, Brian. "Do Judgments Screen Evidence?" Manuscript.

Wedgwood, Ralph (2010). The Moral Evil Demons. In Richard Feldman & Ted A. Warfield (eds.), *Disagreement*. Oxford University Press. 216–246.

White, Roger. "On Treating Oneself and Others as Thermometers." *Episteme* 6, no. 3 (October 2009): 233–250. https://doi.org/10.3366/E1742360009000689.

White, Roger. "You Just Believe that Because…" *Philosophical Perspective* 24, no. 1 (January 2011): 573–615 https://doi.org/10.1111/j.1520-8583.2010.00204.x