Philosophers' Imprint

volume 14, no. 18
june 2014

Disagreement about Disagreement? What Disagreement about Disagreement?

Alex Worsnip

Yale University

Disagreement is a hot topic in epistemology. A fast-growing literature centers around a dispute between the "steadfast" view, on which one may permissibly maintain one's doxastic attitudes even in the light of disagreement with epistemic peers who have all the same evidence, and the "conciliationist" view, on which such disagreement requires a revision of these attitudes.[1] The battle lines here, it might seem, are clear.

On closer inspection, however, there is something odd about the disagreement debate. For pretty much all steadfasters and all conciliationists agree that (i) there are some situations in which you should revise your doxastic attitudes in the face of disagreement with a heretofore-thought epistemic peer; and (ii) there are some such situations in which you should not revise these attitudes, or at least not by much. In terms of the practical advice that the views issue, then, the initial clear divide between the "steadfast" and "conciliatory" views is not so clear as it first appears. Moreover, even pinning down what separates them theoretically can be a tricky matter.

One possible response to this situation open to a theorist from either camp is to try to present the other side's central cases as exceptions to a broader pattern, and to explain what makes them exceptional. That way, it might seem, the theoretical differences between the camps can be preserved — and maybe they will differ over more intermediate cases.

A different approach, and the one I will pursue in this paper, is to develop a view that is moderate all the way down. On this view, there is a more comprehensive and gradual continuum of cases of disagreement, varying from those which call for radical revision of doxastic attitudes to those which call for no or virtually no revision at all — with many gradations in between — differing primarily in degree, not in kind. Those tempted by a view like this are sometimes pessimistic about the prospects for giving a unified account that clearly

1. For the former view, see Kelly (2005, 2010), Van Inwagen (2010), and Titelbaum (forthcoming), amongst others. For the latter view, see Christensen (2007, 2011), Elga (2007), and Feldman (2006), amongst others.

predicts *when* more or less extensive revisions will be called for.[2] By contrast, in this paper I will give an account that aspires to such unity and predictive power.[3] The view I will present thus amounts to a new, moderate theory of how one should respond to disagreement.

The crucial notion for the view I will defend is the "net resilience" of one's estimate of one's own reliability against one's estimate of one's interlocutor's reliability. I will explain what this means and why it matters for how we should respond to disagreement. However, I will also argue that ultimately, when we weaken conciliationism and the steadfast view to account for exception cases, and to make them adequately plausible, they end up converging on the moderate view I present. Much of the seeming disagreement about disagreement is, I will argue, illusory.

I proceed as follows: section 1 addresses the setup and conceptual framework required to present competing views about disagreement, building up to introducing the key notion of net resilience. Section 2 begins the substantive argument by presenting extreme versions of conciliationism and the steadfast view in light of this framework, and giving some simple counterexamples to them. These views are presented not because they are accepted by many philosophers as they stand, but simply as a foil to establish the point that any plausible view will have to concede that conciliation and steadfastness are each appropriate in at least some circumstances. Section 3 answers a theoretical challenge posed by extreme conciliationism to the idea that steadfastness is ever called for, even in extreme cases. Section 4 presents my moderate account in light of the criticisms of the extreme views, and explains how net resilience makes the difference between cases in which significant conciliation is called for and those in which it is not. Section 5 deals with an important objection to my view inspired by some work by Roger White. Finally, sections 6 and 7 argue for the convergence of moderate conciliationist and steadfast positions (respectively) on the moderate view that I favor.

2. See esp. Enoch (2010).

3. Lackey (2010a, 2010b) also presents her view in such a way. See section 4 for the differences between our accounts.

1. A framework for presenting the competing views

One might naïvely think that the question at issue in the disagreement debate is as follows: when you encounter a dispute with an epistemic peer, should you maintain your belief or abandon it, suspending judgment? However, this way of framing things is oversimplified.

First, one's options when one is confronted by a disagreement are not just to maintain belief or suspend. Rather, one can alter one's credence in the proposition at issue by a range of degrees, as is well-appreciated in the literature.

Second, a similar point holds about 'epistemic peerhood,' a term which can be vague or differ in meaning from theorist to theorist, and is consequently capable of being misleading. Discussion of epistemic peerhood sometimes proceeds as if the relevant question is whether one's interlocutor is intelligent or worthy of intellectual respect.[4] This is distorting. What matters when it comes to disagreement is how likely my peer is to be *right*, that is, how *reliable* she is. And I can respect someone's intellectual powers greatly while thinking her actual views very unreliable. While there is a sense in which I should think of such a person as my 'epistemic peer,' it is not the relevant sense for whether it will be rational to conciliate with her. Intelligence is not the same thing as reliability.[5]

4. Compare, especially, Van Inwagen (2010: 23–24).

5. It might be objected that, in appealing to the claim that I can take someone to be unreliable whilst respecting them as having equal intellectual powers to mine, I beg the question against the conciliationist position. According to the conciliationist, this objector says, there is no reason to take someone to be less reliable than me unless I can chalk it up to some difference in intelligence. But this misunderstands the conciliationist. There is no reason why a conciliationist should commit herself to the claim that intelligence is all I can legitimately go on in estimating a would-be peer's reliability. Intelligent people can have a range of other characteristics that make them bad truth-trackers: they can be under the sway of distorting doctrines and ideologies; they can have (perhaps subconscious) tendencies to believe and defend extreme or radical positions for fun; they can lack common sense; they can be psychologically

Given that it is reliability that is at issue, it seems clear that one might not think that others are exactly as reliable as oneself all that often.[6] Perhaps more often, one thinks that one's interlocutor is at least somewhat more or less reliable than oneself; again, there is a whole range of possible estimated levels of reliability.

With this clarified, let us operate with a standard range of possible credences in a proposition p from 0 to 1, where 1 marks certainty in p, 0 marks certainty in not-p, and 0.5 marks lending equal credence to p and not-p.

It's more obvious how to operationalize the notion of reliability in the context of outright belief than in the context of credences. With outright belief, we can just talk of whether someone's beliefs are true or false, and then think of reliability as the propensity for her beliefs to be true. But not so with credences. And in my view, we also shouldn't measure reliability as a doxastic agent's propensity to have a credence close to 1 when a proposition is true and close to 0 when it is false. If we did this, then one would count as more reliable for having more conclusive evidence to go on: two agents could be equally ideal, yet the one with more evidence would count as more reliable. But we might want to treat evidential position and reliability as independent rather than collapsing them.

There are, however, more sophisticated ways to understand reliability in the context of credences.[7] Here is one, just to fix ideas. Let the **ideal credence** be the credence which the agent's evidence supports. Now we can define the relevant notion of **reliability** as the agent's propensity to get close to the ideal credence given her evidence. One way to measure this would be to measure the probability that the agent's credence is closer to the ideal credence than a randomly selected credence is to that ideal credence. But I set these details aside here.

As just mentioned, I am treating reliability and evidential position as independent. For the purposes of this paper, I will confine myself to cases where one takes oneself to share the same evidence with one's interlocutor.[8] A fully general model of how to respond to disagreement would relax this assumption, and allow for uncertainty about the relative evidential positions, either by treating this as a separate variable or by incorporating it into the measure of reliability. I believe that this can be done, in much the same way that I account for uncertainty about one's interlocutor's reliability in this paper. But doing so is not crucial to my project of showing how a moderate view on disagreement can unify the different existing approaches, and nor does this complication make anything I have to say here inaccurate. So, since I do not wish to overcomplicate the account, I propose to leave this extension of the account to future work.

We can now define the **net reliability** in a disagreement to be the difference between your reliability and that of your interlocutor. So, to map this onto one traditional use of 'epistemic peer,' you would be epistemic peers with an interlocutor when the net reliability is set at zero.

You also have estimates of all these reliabilities. I take it that it can sometimes be very non-transparent to you what your credal response to the evidence should be; you can also be aware of a track record that indicates that you do not always get close to the ideal credence. So, I assume that you ought not, usually at least, judge yourself to be

invested in particular views for arbitrary historical reasons. Whilst estimating these things can be hard, and it can be hard to do so in a way that sets aside the possibility that one's judgment is clouded by one's own possession of these characteristics, the same is true of assessing (lack of) intelligence. So, I see no reason why the conciliationist should limit herself to the latter in assigning the reliability of her would-be peers.

6. For a detailed discussion of this point, see King (2012).

7. For example, in terms of calibration; see Lam (2013) for discussion of some possible alternatives.

8. Again, one might think that we rarely share the same evidence for a proposition in the way that the disagreement debate standardly assumes we can (on this point, see esp. King 2012: 253–258). One might think this especially if one follows Williamson (2000: esp. Chs. 8–10; 2007: Ch. 7) in thinking that evidence consists entirely of true or even known propositions.

perfectly reliable. So, you have a **reliability estimate** both for yourself and for your interlocutor.[9]

Now, you can also be more or less confident that each reliability estimate is (roughly) correct — that is, that it approximates the actual reliability of the agent.[10] For example, if you have little to go on with respect to your interlocutor's reliability, you will be relatively unconfident in your estimate of your interlocutor's reliability. This level of confidence in the reliability estimate can be helpfully measured by the **resilience** of the reliability estimate in response to future evidence. The more confident you are in a reliability estimate, the more resilient it is in response to future evidence of reliability, since that evidence will then be relatively insignificant compared to what you already have.[11]

Crucially, having an unresilient estimate of reliability is very different from having a low estimate of reliability. Your reliability estimate is just your best attempt at guessing a reliability. But this could be based on extensive track record information, or on almost nothing. There could easily be two people such that I judge both to be equally reliable when forced to estimate their reliability, but whereby I am much more confident in my judgment with respect to one person than the other. Accumulating track-record data about someone's reliability may serve to increase my estimate of her reliability in some cases, but in others it may serve to raise my second-order confidence in my estimate of her reliability, without actually increasing that estimate.

In the same way, there can be a gap between the resilience of my estimate of my own reliability and that of my estimate of my interlocutor. Call this the **net resilience**, where a positive net resilience indicates a higher resilience of my estimate of my own reliability than my estimate of my interlocutor. Since the resilience of a reliability estimate is not a function of its value, net resilience is likewise not a function of your estimate of the net reliability. I can estimate our reliability at the same

9. In talking of one's estimates of reliability and their resilience, I may cause the reader to wonder whether I really intend my account to operate in terms of one's actual psychological estimate of reliability, or whether it should be framed in terms of the reliability estimates that one's evidence supports. (Thanks to an anonymous referee for raising this point.) I myself am sympathetic to the idea that the requirements governing one's response to disagreement are "wide scope": they tell you that it is required of you that, if you begin with a particular set of credences, reliability estimates and resiliencies, and you encounter a disagreement, then you revise your credences in some particular way. Here the credences and reliability estimates are to be understood as the ones that you *actually* have. But because the 'requires' operator takes wide scope over the whole conditional, one cannot "detach" a requirement to fulfill the consequent of this conditional even when the antecedent is satisfied. So, if one starts with all the wrong credences and reliability estimates, one would not be *required* to have the credence that would result from following the disagreement norm beginning with these out-of-whack estimates. And this is good, because that credence might be one that one ought not to have. In such conditions, following the disagreement norm is not enough to guarantee that one ends up with the credence that one ought to have. Nevertheless, the fault here is with one's original estimates, not with one's failure to observe the disagreement requirement correctly: one has satisfied *that* requirement, and it is possible to satisfy it even when one's original credences and estimates are incorrect. So, although this account uses one's actual reliability estimates, it does not amount to a simple "subjectivization" or "psychologizing" of an account that operates in terms of the reliability estimates one ought to have given the evidence. (For more on wide-scope requirements, see Broome 1999; 2013: esp. ch. 8; Dancy 2000: 60–76. I develop the view with respect to rational belief in Worsnip ms-b.)

   I hope to explore the idea of reading disagreement requirements with wide scope in future work. In this paper, however, I am not relying on it. The account here could easily be altered, for those who reject the wide-scope account, to simply utilize the estimates that one ought to have given the evidence. Any such readers should read my talk of "estimates" of reliability accordingly.

10. Note that such a lack of confidence in your estimate may not reflect any lack of confidence that your estimate is the best possible estimate given the evidence you have. It may simply reflect limitations in how much evidence you have, which make it unlikely that your estimate of reliability approximates the *actual* reliability.

11. For general discussions of the importance of resilience in doxastic life, see Skyrms (1980) and Joyce (2005). Note, however, that here it is not the resilience of your *credence* that we are interested in, but the resilience of your *estimate of your (and your interlocutor's) reliability*. Here is an example to show how the two can come apart: you tell me that in one of your clenched fists is a $100 note, but do not tell me which. My credence that the $100 is in your left hand is 0.5. The resilience of my *credence* is low: a small amount of evidence could easily change it. But I take myself to be very reliable in getting close to the ideal credence (which is a notion sensitive to my evidential position) in cases like this, and the resilience of this estimate of my reliability is high.

level but have a much higher resilience for my estimate of myself than for my estimate of you.

One might be tempted to think of the net resilience as tracking my level of confidence that I am just as reliable as my peer. But this is not quite right. Suppose that my estimate of my own reliability and my estimate of my interlocutor's reliability are equal and, moreover, very resilient to an equal degree. In this case the net resilience is zero, and I am confident that we are equally reliable. Now compare a situation where both my estimate of my own reliability and my estimate of my interlocutor's reliability are, though still equal, very unresilient to an equal degree. In such a case, my net resilience is still zero, but I am not at all confident that I am as reliable as my peer. So, in both cases the net resilience is zero, but in one case I am much more confident that we are equally reliable than in the other.

What is needed for the net resilience to be high is that there be a large *gap* between the resilience of my estimate of myself and the resilience of my estimate of my peer. According to the account I will ultimately defend, it is the net resilience which matters — not my confidence that my peer and I are equally reliable.[12] So, to know how you should respond to disagreement, we need to know your individual reliability estimates and their resiliences — not just your level of confidence in the claim that the two of you are equally reliable.[13]

12. Thanks to an anonymous referee for pushing me to clarify this point.

13. An anonymous referee worried that your belief in peerhood might be hard to "factorize" into two reliability estimates in this way. But although it is always hard to attribute very precise credences to agents, I do not think it is implausible that rational agents have factorized judgments about reliability in disagreement cases. Suppose we have an agent who is very uncertain about whether she and her peer are equally reliable. It seems that, if she is responding to (what she takes to be) evidence, there should be a determinate fact of the matter as to whether her uncertainty is due to (i) low resilience for both reliability estimates or (ii) low resilience for just one of those reliability estimates. Of course, she probably wouldn't put it this way herself. But we could imagine asking her in more colloquial terms: "so, you're not too sure that you and Mary are equally reliable. Is that because you're not sure how reliable she is, or because you're not sure how reliable either of you are?" If she's actually basing her attitudes on the relevant evidence, it seems that there should be a determinate answer to that question, so that her reliability judgments are "factorizable." Likewise, someone with a high confidence that she and her peer are equally reliable needs to be justified in having a high resilience for both parties in order to be justified in having this high confidence.

Nevertheless, it is true that when both estimates are very resilient (as opposed to merely equally resilient), one will be confident that one is about as reliable as one's peer. This highlights an ambiguity in the existing disagreement literature. When it refers to interlocutors that you "(justifiably) believe to be your epistemic peer," or similar, it equivocates.[14] This could mean that, when forced to issue an estimate of their reliability, you (justifiably) estimate that reliability at a level very close to your own estimate of your own reliability, based on what you have to go on. Or it could demand more: that the estimates, in addition to being equal, both be (justifiably) highly resilient, so that you (justifiably) believe the proposition expressed by 'our actual reliabilities are approximately equal.'

This second possibility makes epistemic peerhood much harder to attain than the first does. In particular, one can (justifiably) estimate one's own reliability at precisely the same level as one's peer's, but not (justifiably) believe that the two actual reliabilities are approximately equal, if one or more of the reliability estimates is unresilient. This means that even though one estimates both reliabilities at the same level, one is not at all confident that this estimate reflects the actual reliabilities. Indeed, one might (justifiably) think it positively unlikely that the reliabilities are actually the same, whilst (justifiably) estimating them at the same level based on what one has to go on.

2. The extreme views

Now we're in a position to consider some possible views. Let's start with what I'll call "extreme conciliationism," which is a development of what is sometimes called the "equal weight" view:

14. *E. g.* Elga (2007: 484), Lackey (2010a: 303–4), and Enoch (2010: 971–2), who misses this ambiguity in his otherwise comprehensive disambiguation of 'peer disagreement'.

> **Extreme conciliationism**. Whenever you encounter a disagreement with someone who possesses the same evidence as you, you are required to adjust your credence in the proposition at issue in proportion with your prior net reliability estimate and your relative credences. In particular, if your net reliability estimate is 0, and your credence in p is equal to your interlocutor's credence in not-p, then you are required to adjust your credence to 0.5.[15]

Extreme conciliationism is not even slightly plausible. Consider the following case:

> **The lizards**. At a philosophy conference in Washington DC, I meet David.[16] Considering myself to be a roughly average epistemic agent amongst philosophers, and David to be a randomly sampled philosopher, I am initially inclined to make a net reliability estimate of 0, though of course my net resilience is high, since I have very little to go on about David so far. After a bit of idle chatter about the weather and the conference program, I mention that on my off-day I am planning to go and see the White House. "Don't go there!" exclaims David. "The US government is run by a sinister race of disguised lizards. All the major governments of the world are."

David and I disagree about whether the major governments of the world are run by a sinister race of disguised lizards (henceforth, **LIZARDS**). And, prior to our conversation about lizards, I assigned David approximately the same level of reliability that I assigned

15. The closest to an advocate of this extreme view in the literature is Elga (2007).

16. So named for David Icke, a former football player and sports journalist from England who espouses roughly the same theory as my imagined interlocutor. See http://en.wikipedia.org/wiki/David_Icke

myself. Suppose David is just as convinced that LIZARDS is true as I am convinced that LIZARDS is false, and that there are no major differences in the evidence available to myself and David. Am I now required to increase my confidence in LIZARDS to 0.5? Of course not. Rather, what I will and ought to do is radically downgrade my estimation of David's reliability, at least with regard to this domain of inquiry. I will not adjust my credence in LIZARDS to any non-trivial degree.

Consider next what I will call "extreme steadfastness:"

> **Extreme steadfastness.** Encountering disagreement, at least with anyone who you do not estimate as more epistemically reliable than you, is never in itself a reason to adjust your credences.[17]

Extreme steadfastness is also not even slightly plausible. Consider the following kind of case, familiar from the literature:[18]

> **Dinner check.** I go out for dinner with seventeen friends. At the end of the meal, my friend Cat and I both calculate what our third friend, Alfonso, owes. It is a slightly tricky calculation, but not one beyond either of our powers. I take Cat to be equally reliable as me at mental arithmetic, on the basis of numerous similar past occasions. I come up with $21.74, and form credence 0.9 in the proposition that Alfonso owes $21.74. Next I discover that Cat has credence 0.9 that the amount Alfonso owes is $22.74.

Quite obviously, I should revise down my credence in the proposition that Alfonso owes $21.74. More generally, the idea that I could justifiably maintain the same credences about some proposition p regardless of whether *every person I know that I do not judge to be my*

17. The closest to an advocate of this extreme view in the literature is the early Kelly (2005: see esp. his formulation of his view on p. 170).

18. See, *e.g.*, Christensen (2007: 193).

epistemic superior agrees, orwhether *no people I know that I do not judge to be my epistemic superior* agree, is incredible. Clearly, such variation can serve as relevant evidence in determining whether p is true.[19]

3. The circularity charge

Well, those are the intuitions. But extreme conciliationism does at least pose us a challenge. It goes like this: revising my evaluation of David's reliability in **the lizards** seems in some way problematically circular. David and I are disagreeing about LIZARDS, and I am resisting revising my credence on the grounds that David is unreliable. But my assessment of his reliability in turn depends on me assuming that he is wrong to believe LIZARDS. So, how could this be grounds for remaining steadfast in my very low credence for LIZARDS?

I think this is the wrong way to think about things. I am not resisting revising my credence "on the grounds" that David is unreliable, in some way that rests upon the assumption that my credence is correct. Rather, I am considering two possible courses of action in response to our disagreement: revising my credence, and revising my estimate of David's reliability.[20] This takes place in response to a kind of mismatch: given our disagreement, it seems that I ought not both have a low credence in LIZARDS *and* attribute high reliability to David. Holding either my credence or my reliability-attribution constant, it will seem that the disagreement provides evidential grounds for revising the other. It's seriously misleading to say that, in resisting revising my credence, I am doing so "on the grounds" that David is unreliable.

If I were doing something problematically circular here, it's hard to see why the alternative response to the situation — revising my credence rather than my reliability-estimate — would not also be circular. After all, it is just the other choice that's available to get rid of the mismatch. Just as the extreme conciliationist tries to demand that I hold my antecedent reliability estimate constant in the face of disagreement, we could demand that I hold my antecedent credence constant in the face of disagreement. Then we could put the following parallel circularity objection to me if I conciliate: "you are resisting revising your reliability estimate on the grounds of your (now) higher credence in LIZARDS. But your higher credence in LIZARDS in turn depends upon thinking that David is reliable. So, your reasoning is circular!"

If this circularity objection is a bad objection to extreme conciliationism, the extreme conciliationist's circularity objection to alternative views is likewise bad. Thus, I see no reason to accept the extreme conciliationist's view that I must hold my credence in my interlocutor's reliability fixed, and revise my credences in the proposition at issue in light of the former. I can be just as free to revise the latter.

Now, actual conciliationists, unlike extreme conciliationists, usually admit that in cases like **the lizards**, I may permissibly downgrade my

19. In his early (but not his later) work, Kelly (2005: 182–3) appeared to deny this, affirming that only the arguments bearing directly on the truth of some proposition p count as relevant evidence, and that what others think — as an "empirical and contingent fact" — does not. (Oddly, Kelly allows in this same article [2005: 173–74] that a disagreement with someone who is your epistemic superior *can* give you evidence that calls for a revision in your credence, even though his argument could easily be adapted to cover disagreement with epistemic superiors.) Kelly makes this seem more plausible by getting us to imagine two worlds such that the arguments that we are aware of with respect to p's truth are identical across the two worlds, but one world exhibits great consensus about p and the great disagreement. In doing so, however, Kelly effectively stipulates that, in the comparison of the two cases, the difference between consensus and disagreement is mere noise: it doesn't track any difference in the facts. The problem is that a) on the assumption that one's peers tend to be reliable to some extent, their verdicts are not usually mere noise in the way Kelly effectively stipulates; and b) even if this were the case, and the disagreement of one's peers were simply misleading evidence, one would not, as an actual party to the disagreement, be in a position to know this. Indeed, something is misleading only if you don't know that it's misleading. That's why even misleading evidence can call for revision of doxastic attitudes.

20. One might wonder whether there is a third option: maintaining my current credence without downgrading my estimate of David's reliability, chalking it up to an uncharacteristic error on his part. (Thanks to an anonymous referee for raising this point.) In section 4, I will argue that in cases like that of David, one *is* positively warranted in downgrading one's estimate of one's interlocutor's reliability. In section 7, I will take on the proposal that there are other cases where one is justified in remaining steadfast without being warranted in downgrading one's interlocutor, and argue against it.

estimate of David's reliability. Such actual conciliationists nevertheless sometimes want to appeal the circularity objection, often putting the point in terms of a need for my reasons for dismissing my interlocutor to be "independent" of the dispute at hand.[21] *Prima facie*, however, it seems that my decision to downgrade my estimate of David's reliability is *not* independent of the dispute at hand: I have downgraded David in response to the craziness of his claim that LIZARDS is true. But there may be a more subtle notion of independence at work, which counts my reasons for downgrading David as suitably independent, while still counting other cases of downgrading as objectionably non-independent. In section 6, I'll return to this question in much greater detail, in the course of arguing that the more moderate conciliationist view converges with my own.

4. Explaining what makes the difference

I've just argued that, from the point of view of circularity, there is nothing necessarily more objectionable about revising one's reliability-estimate than there is about revising one's credence. But this leaves open which of these options I should in fact take in particular cases. Intuitively, I find it clear that in the case of **the lizards**, it is more reasonable to choose to revise my reliability-attribution than to revise my credence. Conversely, in **dinner check**, the right reaction is to revise down my credence and maintain my reliability estimate. So, whether I should revise my credence or my reliability estimate depends on the particular case. Call this view the **boringly moderate account**. I think that the boringly moderate account is true. But it also seems to me that we should be unsatisfied with simply stating the boringly moderate account as it stands. We want an explanation of what makes the difference between the cases, and we want a more generalized account of when I should revise my reliability estimate, and when my credence.[22] This is what I'll give shortly.

First, an existing proposal from Jennifer Lackey: the difference is that, whereas in **dinner check**, you give some antecedent credence to the chance you might be wrong, in **the lizards**, you (justifiably) give very little antecedent credence to this.[23] There are some things that are beyond the pale, and LIZARDS is one of them; that's what makes the difference.

This cannot, I think, be right. My high confidence that LIZARDS is false serves as a high baseline from which to start my revisions, but it does not on its own explain why there is little to no pressure to move away from this baseline to at least some degree. Perhaps, if I am very confident to start with, my *final* credence will be higher than it would have been otherwise; but *the amount I reduce my credence by* may not be. It's telling here that Lackey's view finds more natural expression in terms of binary belief than in terms of credence.[24] For when it's binary belief that's at issue, a higher starting baseline of confidence could explain why disagreement does not call for the all-out abandonment of the belief, even if that confidence should be reduced substantially. Nevertheless, this does not mean that the high baseline is reducing the amount of pressure to revise the confidence down from that baseline.

Indeed, there can be situations in which I am very confident about something, and even justifiably so, but encountering disagreement warrants reducing this confidence (after all, the disagreement is itself

21. See esp. Christensen (2011).

22. Enoch (2010: 992–95) endorses something similar to the boringly moderate account — calling it the "common-sense view" — but rejects the idea that anything general can be said about what determines the balance between revising one's credence and revising one's reliability estimate. Here, I attempt to be more ambitious, and thereby show that something general can be said about this.

23. Lackey (2010a: 306–8, and esp. 316–9). See also Elga's (2007: 483) claim that his conciliatory view doesn't apply "outside an appropriate range," namely when you find your disputant's contentions "insane" (*ibid.*: 491). It's unclear why, if this can prevent one entirely from having to conciliate for Elga, finding one's disputant's contentions, say, somewhat-less-than-insane-but-still-surprising-and-odd doesn't even temper the need to give her view equal weight.

24. See Lackey (2010a: 310). Lackey considers an objection to her account based on its use of binary belief at (2010b: 282–3), but this objection is not the same as mine and so her response does not address my worry.

new evidence).[25] Roughly speaking, these will be situations where I expect not to encounter disagreement about some matter because I take it to be uncontroversial, yet my being so confident is itself conditional on not encountering such disagreement. It is perfectly coherent to simultaneously have a very high credence in a proposition, yet to have a significantly lower credence in it conditional on encountering a disagreement, provided that one thinks it relatively unlikely that one will encounter disagreement. And there seems no in-principle bar to all of these attitudes being strongly justified, even if the claim that one will not encounter disagreement turns out ultimately to be false. So, the level of one's first-order justified credence cannot directly predict the amount of revision of it that is called for by encountering disagreement.

What really explains the amount of revision called for, I suggest, is my net resilience. In **dinner check**, I have a low net resilience, because I have a lot to go on in estimating Cat's reliability. In **the lizards**, conversely, I have a high net resilience. I attribute a high level of reliability to myself in the circumstances — this is not the kind of case where it seems difficult to work out what the evidence supports — and I am very confident in this high attribution of reliability. Conversely, while I attribute to my interlocutor high reliability — there's no reason to think that he's a lunatic antecedent to the dispute — I am not very confident in this high attribution of reliability, since I have very little to go on in arriving at it. Given that my estimate of my own reliability is fairly resiliently set at a high point, but my estimate of my interlocutor's reliability is much more sensitive to possible evidential updates, my net resilience will be high.

Why does the net resilience make the difference that it does? Suppose that you estimate your interlocutor's reliability and your own roughly equally at some level above 0.5. Then, a disagreement can be evidence that you estimated at least one of these reliabilities too high. After all, both you and your interlocutor have disagreed with a seemingly reliable source, and that is at least some evidence against both of your reliabilities. This point is recognized on all sides. But conciliationists have argued that, since your estimates of your own reliability and that of your interlocutor are equal, the disagreement cannot be evidence against either disputant's reliability more than the other.[26] So, even if you should downgrade your estimate of your disputant's reliability, you should do the same for yourself; so, you are not licensed in resisting revising your credence concerning the matter under dispute.

This has some initial plausibility. And this reasoning does indeed hold as long as the net resilience is zero (regardless of the absolute level of the resiliences – as long as they are the same). But it is an overgeneralization to think that it applies to any case where merely the estimated *reliabilities* are the same. This overlooks the importance of resilience.

Let me explain. First, note that if your estimate of your own reliability is highly resilient, then quite a lot of evidence is required to shake it. Conversely, less evidence is required to shake a less resilient estimate of your interlocutor's reliability. Second, note also that the existence of the disagreement provides different evidence relative to which estimate we are considering it in relation to. The evidence against the accuracy of your estimate of your own reliability is that you have disagreed with someone whom you estimate to be highly reliable based on what little you have to go on. Conversely, the evidence against the estimate of your interlocutor's reliability is that he or she has disagreed with someone whom you estimate to be highly reliable based on extensive evidence and track-record data (namely, you!). The result of all of this is that, when the net resilience is high, so that your estimate of your own reliability is much more resilient than that of your interlocutor, the disagreement can provide on-balance

25. For a compelling example, see Christensen (2007: 200).

26. See, *e.g.*, Christensen (2007: 196–98), and Elga (2007: 487), who calls the idea that a disagreement can provide evidence that you are a better evaluator than your interlocutor "absurd." Though, compare Elga (2007: 491).

evidence that your estimate of your interlocutor's reliability is too high. Indeed, this can be shown mathematically.[27]

Since the net resilience makes a difference to when I can downgrade my estimate of an interlocutor's reliability, it also makes a difference to when I can be steadfast in my credence. When the net resilience is high, and I am permitted to downgrade my estimate of my interlocutor's reliability to a great extent — as in **the lizards** — this eliminates the need to downgrade my credence to any significant extent. However, when my net resilience is low, and I am not permitted to downgrade my estimate of my interlocutor's reliability to any great extent — as in **dinner check** — I will have to downgrade my credence significantly.

These are extreme cases, however. In the bulk of actual cases of disagreement, the net resilience will be more intermediate. The net resilience is almost always going to be some positive non-zero value, just because one typically has much better evidence regarding one's own track record than one does regarding one's interlocutor.[28] But the net resilience will not usually be as high as it is in **the lizards**.

I suggest that in such intermediate cases, the thing to do is to revise *both* my credence in the proposition at hand, and my estimate of my interlocutor's reliability. In this way, I can find a sort of equilibrium.[29] A reduction of either my credence or my reliability estimate tempers the need to reduce the other. If my credences are extremely fine-grained, I may even do some tiny revising of my credence in LIZARDS in **the lizards**, and alter my estimation of Cat's reliability ever-so-slightly in **dinner check** (though clearly, I should reverse this change in the latter case if she subsequently turns out to be right). But not much. The balance of whether I should make more of a revision in my credence in the proposition, or in my estimation of my interlocutor's reliability, depends on my net resilience.

5. White's challenge

Of the existing work on disagreement, the paper which comes closest to discussing the view I have suggested here is White (2009: 247–249). There, White makes some remarks that may provide some resources for an extreme conciliationist to push back against the view. So, it is worth considering the challenge that White issues.

White agrees that what I have called net resilience is of significance in situations of disagreement. However, he thinks that its significance is limited to its effect on your judgment of your peer's reliability.

27. Here is an illustration, adapted from White (2009: 247–48). (Despite the helpfulness of White's example for my purposes here, White himself denies a crucial part of my account of the significance of resilience, as I will explain in section 5.) Suppose that you know for certain that you are 0.9 reliable on some matter. As for your interlocutor, you know she is either .85, .9 or .95 reliable, but have no idea which. So, you estimate both reliabilities as 0.9, but your resilience in your estimate of your own reliability is much higher than that of your estimate of your interlocutor. Suppose a disagreement now occurs. Since you know that you are 0.9 reliable, you can reason as follows: "Conditional on her being 95% reliable, the antecedent probability that she would disagree with me (where I am .9 reliable) was .14. Conditional on her being 90% reliable, the antecedent probability that she would disagree with me (still 90% reliable) was .18. Conditional on her being 85% reliable, the antecedent probability that she would disagree with me (still 90% reliable) was .22. So, the fact that we disagreed alters the balance of evidence between the possible reliabilities for my interlocutor, lending most weight to the 85% possibility, then the 90% possibility, then the 95% possibility. So, I have gained some evidence that should shift my estimate of her reliability down from its original level of 0.9." The argument generalizes for any case where there is a difference in resilience between the estimates of the two reliabilities and where those reliability estimates exceed 0.5. When there is a disagreement in such a case, one gains evidence that should shift the less resilient reliability estimate downwards. The evidence provided by the disagreement is not neutral between that reliability estimate being too low and its being too high.

28. Of course, there are unusual cases where the net resilience is actually negative, such that your estimate of your interlocutor's reliability is more resilient than that of your own. In such unusual cases, the disagreement would actually provide reason to downgrade your estimate of your own reliability, and give your interlocutor's judgment *more* weight than your own.

29. Note that this view does not require any kind of problematic attempt to get outside one's own internal perspective or belief system, a worry that Enoch (2010: 961–65) has about conciliationist views. The incoherence in maintaining a high estimate of my interlocutor's reliability and maintaining a high credence in the proposition under dispute is an entirely internal one that can be felt acutely from a first-person perspective. (I think a similar defense may work for more extreme conciliationist views, though as I have made clear, I reject these views for other reasons.)

Specifically, he thinks that if you encounter disagreement with someone who you judge to be equally reliable as yourself, but your estimate of your own reliability is more resilient than that of your interlocutor, then you are warranted in downgrading your estimate of your interlocutor's reliability. Here we agree. But White denies that this warrants you in tempering your reduction of credence in the matter under dispute.[30] This reduction of credence, he thinks, depends on your prior estimates of reliability, but not on their resilience. White takes this to be a straightforward result of Bayesian conditionalization: given your equal estimate of the two reliabilities, you should take it that either is equally likely to err. So, if, for example, you and your interlocutor have pre-disagreement credences for some proposition p that are equally distant from the midpoint of 0.5,[31] your prior credence for p conditional on a disagreement should be 0.5. So, upon learning of a disagreement, you should revise your credence in p to 0.5. The resilience of the reliability estimates has no effect here.

Let's begin by noting the intuitive costs of this view. First, most simply, it entails that in **the lizards** — if I really have estimated David to be as reliable as me (just with lower resilience in this estimate) — the right response to the disagreement with David is to increase my credence in LIZARDS to 0.5. That is, on its own, still an incredible view. But moreover, White's concession that I *should* downgrade my estimate of my interlocutor's reliability in such a case yields a further very odd result. Since the resilience of my own reliability is high, my estimate of my own reliability is largely unaffected. Nevertheless, my estimate of David's reliability is brought down by White's lights, potentially to a great degree. Suppose that I follow White's advice and downgrade my estimate of David's reliability, but nevertheless increase my credence in LIZARDS to 0.5. I'm now left with the following set of judgments at this point in time:

- My initial verdict on this matter was that the world is not run by disguised lizards.
- David's initial verdict was that the world is run by disguised lizards.
- I estimate my initial verdicts on matters like these to be much more reliable than David's.
- Nevertheless, I'm no more confident that the world is not run by disguised lizards than that it is.

This set of beliefs seems bizarre at best and incoherent at worst. If your verdicts on matters like these are more reliable than David' — and your best estimate, based on your latest assessment of the evidence, is that they are — surely by your own lights you have evidence favoring the hypothesis that the world is not run by disguised lizards. And it is evidence which, it seems, has not been factored into your credence. It's not like there is some other strong reason to prefer the lizards hypothesis which is counteracting it. This suggests that intuitively, White is wrong to say that the downgrade in the estimate of David's reliability should not temper the downgrade in credence for the proposition that the world is not run by disguised lizards.

Does this intuitive objection require the rejection of Bayesian conditionalization? No. Note that there are two basic steps in White's derivation of his result. First, there is the claim that given an equal estimate of my own reliability and that of my interlocutor, my prior credence in p conditional on the disagreement should be 0.5. Second, there is the simple move of Bayesian conditionalization, namely that given that prior credence, upon learning of a disagreement my credence in p should move to 0.5. Now, if we try to block the second

30. By contrast, in his original presentation of the equal weight view, Elga (2007: 486–8) thought that it was precisely because you're *not* permitted to downgrade your estimate of your interlocutor's reliability that you have to give her opinion equal weight: indeed, that is his central argument for the equal weight view. So, Elga seems to implicitly agree with me, against White, that *were* this downgrade permissible, the need to conciliate on your credence would be tempered. See also Weatherson (ms.: 9).

31. I'll continue to work with this simplifying assumption throughout this section. All the arguments generalize easily once the assumption is relaxed.

step, then obviously we violate conditionalization.[32] But not if we try to block the first step. Can we do that?

Yes: we should claim that, when I have an equal estimate of our two reliabilities but a much higher resilience in my own estimate, my prior credence in p conditional on the disagreement should *not* be 0.5. Why? Because, in such cases, the disagreement is evidence that the initial estimate I made of my interlocutor's reliability is out of whack with my interlocutor's actual reliability. Crucially, recall that this is something which White concedes in admitting that the disagreement should lead me to revise my estimate of my interlocutor's reliability down. If that's right, then it seems I should not calculate my new credence based on the old, faulty, reliability estimate.[33]

This isn't to deny that generally speaking, there's a connection between reliability estimates and credences. If I estimate my interlocutor's reliability at 0.9, for example, then generally speaking I should have a credence of 0.9 that she'll answer correctly, as White points out (2009: 234). So, then, for example, if she just tells me her answer to a question that I've never considered, then I should (*ceteris paribus*) form a credence of 0.9 in her answer. But there's always going to be a caveat to this connection between reliability estimate and credence: it holds only barring any new evidence that bears on the reliability estimate. So, if I estimate your reliability at answering math questions at 90%, but then you act out your answer by singing it to the tune of Verdi's *Requiem* while throwing a bucket of anchovies over your own head, then I have gained new evidence that you may have gone mad — in which case your reliability may be worse than I first thought — and consequently I should not form credence 0.9 in your answer.[34] This, obviously, is no violation of conditionalization: while my conditional credence for p given your answering p is 0.9, my conditional credence for p given your answering p *and* singing and throwing anchovies is somewhat lower.

In the case of disagreement, given a resilient estimate of my own reliability and an equal, but less resilient, estimate of your reliability, the very occurrence of a disagreement is evidence that you are less reliable than first thought (to reiterate, White concedes this). And so, just as we should not say that I am required to calculate my credence for p based on my old reliability estimate in the *Requiem*-and-anchovies case, we should not say that I am required to calculate my credence for p based on my old reliability estimate here. Again, this requires no violation of conditionalization. My credence for p conditional on a disagreement between us should not be 0.5, even though I currently estimate our reliability as equal, because I know in advance that were that disagreement to occur, that would be evidence that we are not in fact equally reliable. This allows us to save the view advanced here from White's objection.

The point here mirrors a now well-established point in the literature on conditionals. Adams' Thesis, roughly stated, is that the probability of a conditional is equal to the probability of its consequent conditional on its antecedent.[35] This thesis initially seems plausible, but there are counterexamples to it that take a specific form.[36] Here is one example.[37]

32. Lackey (2010a: 314) seems to take it that the first step is impossible to block; consequently, she denies the second step. Thus, her account *does* seem to violate conditionalization. Here, I am trying to show that, contrary to what both White and Lackey assume, the first step can be blocked so as to avoid this result.

33. Again (c.f. fn. 10), I don't mean that you've received evidence that your old estimate was an incorrect response to your previous evidence. You've just received new evidence that calls for an update of that estimate by suggesting that it is out of whack with the *actual* reliability of your interlocutor.

34. Indeed, White thinks (2009: 241) that when one gains evidence that a source is inaccurate on a given occasion, this can be reason not to line one's credence up with one's estimate of its general reliability. Rightly, he doesn't take *this* to be a violation of conditionalization! So, it's unclear why this wouldn't also apply when one gains an even stronger kind of evidence: that it is in fact unreliable more generally, and that one's original estimate of its reliability was mistaken.

35. See Adams (1965).

36. See McGee (2000) for the original counterexample, and Kaufmann (2004) for a systematic account of what generates such counterexamples.

37. I learned of this example from Steve Yablo. I have adapted it somewhat here.

Say I have a coin, and I have a 0.99 credence that the coin is fair. Now, consider the following indicative conditional: *If the coin comes up heads 1,000 times in a row, it will be a huge coincidence (that it came up heads 1,000 times in a row)*. Call this conditional proposition C. If the coin is fair, C is true. So, given that I have 0.99 credence that the coin is fair, I will assign at least a 0.99 credence to C. This is not a mistake: it would be a huge coincidence for a fair coin to come up heads 1,000 times, and I have no reason to doubt my high credence that the coin is fair.

However, in the (by my lights) unlikely event that the coin were *actually* to come up heads 1,000 times is a row, I would *then* have reason to become much less confident that the coin is fair. And if the coin is not fair, C may be false: if the coin is sufficiently biased, it isn't a big coincidence that it came up heads 1,000 times in a row. So, conditional on the coin coming up heads 1,000 times in a row, I assign a credence much lower than 0.99 to the proposition that it will be a huge coincidence that it came up 1,000 times in a row. In other words, my credence in the consequent of C conditional on its antecedent is not equal to my credence in C. So this is a counterexample to Adams' Thesis.

What generates the counterexample is the fact that when one learns the truth of the antecedent of the conditional, one gains evidence that forces one to rethink the assumptions on which one assigned a high credence to the conditional (in this case, that the coin is fair). The right response to this case, however, is *not* to hang on to Adams' Thesis by rejecting conditionalization and holding that the probability of the consequent of C conditional on the antecedent of C is 0.99, but that upon actually learning of the truth of the antecedent, one should not update by conditionalization. Rather, it is to hang on to conditionalization by rejecting Adams' Thesis, holding that even though one has credence 0.99 that the coin is fair, and this proposition just entails that C is true, this should only take one to a 0.99 credence in C, and not to a credence of 0.99 for C's consequent conditional on its antecedent.

The exact same lesson applies in the disagreement case. Here my existing estimates of our reliabilities are analogous to my credence in C. Upon learning of an actual disagreement, though, I gain evidence that undermines those estimates, if the net resilience is a positive non-zero value. The lesson, however, is not that I should therefore violate conditionalization. Rather, it is that my credence for p conditional on a disagreement should not be calculated using my existing reliability estimates, but rather by using my estimates of reliability conditional on a disagreement. This doesn't mean I'm not justified in having those estimates *now*, just as I can still be justified having high credence in C now. But they will be undermined if an actual disagreement occurs, just as one's credence in C will be undermined if one learns the truth of its antecedent.

One might nevertheless worry that my situation in cases of disagreement is weirdly unstable. If I'm going to decrease my estimate of your reliability as soon as we disagree, how am I really treating you as equally reliable — as my peer — even now? But remember that disagreement between the two of us is not *inevitable*. Indeed, if I take us both to be highly reliable, I positively expect us to agree. Like the possibility that you'll perform your answer while throwing anchovies, the future event that your answer will disagree with mine is one that I assign a non-zero but non-one credence to, and it's only if it obtains that I'll downgrade you. Moreover, as White points out (2009: 249), if my estimate of your reliability is less resilient than that of my own, then I should also be willing to *increase* my estimate of your reliability — to higher than mine! — in the event that our answers agree. And I'll also use my present estimate of your reliability to respond to verdicts of yours about matters about which I have yet to form an opinion, or when I lack access to the evidential base that you have. So, the fact that my estimate of your reliability would be downgraded in response to disagreement does not make that reliability-estimate — and its status as equal with mine — empty or toothless.

6. What disagreement about disagreement I: conciliationism

As I've said, real-life philosophers do not usually accept extreme conciliationism or extreme steadfastness. They accept watered-down versions. So, once we take account of these modifications, is there anything ultimately dividing them? And, most importantly — having dealt with the White-type objection — is there anything to stop both sides from endorsing the resilience-based view?

Suppose we simply define conciliationism as the view that one is typically required to reduce one's credence in a proposition in the light of disagreement, and steadfastness as the view that one is typically permitted to downgrade one's estimate of one's interlocutor's reliability in the face of disagreement, to avoid giving equal weight to the two views under dispute. As my account shows, both of these claims can be true. All that is required is that in most situations, the net resilience be intermediate, such that both a revision in credence and a revision in one's estimate of the interlocutor's reliability are called for. Under these definitions, then, they are not in disagreement with each other. What, then, is left to separate them?

Christensen (2011: 1–2) thinks that what divides steadfast and conciliatory views is whether they allow us, in revising the estimates of our interlocutors' reliability that we use to determine how to respond to disagreement, to rely on our reasoning concerning the proposition under dispute. According to Christensen, conciliationist views require that the reasons for downgrading one's estimate of one's interlocutor's reliability, and for correspondingly resisting a full-scale revision of one's credence, be independent of the dispute at hand.[38]

Earlier, in section 3, I argued that construed in the most naïve way, this independence requirement is dubious. In **the lizards**, it seems that I am permitted to downgrade my estimate of David's reliability. And doing so seems to be a direct response to his crazy view, LIZARDS. In this respect I may seem to come down on the steadfast side of Christensen's divide.

38. Kelly (2013), a steadfast theorist, accepts this way of dividing the views.

However, Christensen himself wants to avoid the extreme results that one is not permitted to downgrade people like David, and that one is required to conciliate significantly with them. So perhaps he can claim the reasons for downgrading David actually *are* independent of the dispute at hand. Indeed, when Christensen himself addresses cases in which it does seem intuitively as though one is permitted to downgrade one's estimate of an interlocutor's reliability, he appeals to something quite similar to net resilience — without putting the point in these terms. Christensen considers a case where he and a friend disagree on the answer to a calculation question that seems obvious and easy. As he writes:

> "I can eliminate (via personal information) many of the ways in which I could have failed to use a reliable method […] But I cannot eliminate analogous possibilities for my friend. So it's likely that she did not sincerely announce a belief that was formed by a highly reliable method." (Christensen 2011: 10)

In the case that Christensen is describing, this likelihood that the friend did not use a reliable method is something that is arrived at *after* the fact of disagreement. That, Christensen thinks, is the best explanation for why she has given what seems like a crazy answer to the question, assuming that she was being sincere.[39] Remember that, in the case being considered, one judges oneself equally reliable as one's friend coming into the dispute. So, in issuing the answer that he does, Christensen is allowing for a downgrade of his estimate of his friend's reliability (in the circumstances) in response to a disagreement. What licenses such a downgrade? Judging by the above passage, it seems to

39. I have been assuming that we are interested in cases where one knows that there is an actual disagreement, and one's interlocutor is not joking or lying. Like Christensen, I am happy to allow that when it is a serious possibility that one's interlocutor is not being sincere, this can also be a reason to refrain from immediate conciliation. Note, however, that in such a case, one would actually *not* downgrade one's estimate of one's interlocutor's reliability, if by reliability we mean the reliability of their *credences*, as opposed to their *utterances*.

be the fact that he is more *confident* that he is reliable than he is that his friend is reliable. In other words, his net resilience is high.[40]

So, I think that Christensen's reasons for resisting a revision of one's credences in such cases ultimately come down to net resilience. The question we now face is whether such reasons are properly described as independent of the dispute under consideration. Christensen might argue here as follows: it's a precondition of downgrading one's interlocutor that one antecedently have a lower resilience for one's estimate of her reliability than one has for one's own reliability. And these antecedent resiliences are independent of the matter under dispute. So, the reasoning is properly described as independent.[41]

In at least some sense, however, this is misleading. In the most paradigmatic case where one resists revising one's credence in a disagreement on grounds *independent* of the dispute, one does so because one antecedently thought one's interlocutor less reliable than oneself. In such a case, one already thinks — before encountering the disagreement — that one's interlocutor is less reliable on these matters than oneself. In the cases at hand, however, things are different. One does not antecedently think one's interlocutor less reliable than oneself, and it is the *disagreement itself* which causes one to downgrade one's estimate of one's interlocutor's reliability. Were it not for the reasoning that produced the disagreement, one would still be estimating one's interlocutor's reliability as equal to one's own. Granted, it is *also* a precondition of this downgrade that one have a high net resilience, and that net resilience might itself be disagreement-independent. But the disagreement itself — and thus the reasoning the produced it — is nevertheless necessary for the downgrade, which is what justifies resisting revising one's credence. So, there is at least a good sense in which one does not resist revising one's credence on fully independent grounds in such cases.

Nevertheless, there is an important point behind Christensen's attempt to enforce an independence requirement. Armed with the language of resilience, we can express more clearly. The point is this: one cannot refuse to revise down one's credence in p, or indeed revise one's estimate of one's interlocutor's reliability down, simply on the grounds that one's credence in p is high — that one finds the denial of the claim too implausible. Rather, one must appeal to one's net resilience.[42]

Here again we see that an inattention to the distinction between reliability and resilience obscures the debate over disagreement. Clearly, a refusal to revise one's credence on the basis of one's high credence for p does not count as independent. And clearly a refusal to revise one's credence on the basis of an antecedently low estimate of one's interlocutor's reliability *does* count as independent. But what about a refusal based on a *non-antecedent* downgraded estimate of one's interlocutor's reliability, formed in response to the disagreement, but due to an antecedently low net resilience? Because the debate has not even been framed so as to clearly distinguish this third possibility, it is not obviously determinate how the major parties to the debate answer this question. But it may well be that while conciliationists are thinking of this sort of reasoning as satisfying the independence requirement, steadfast theorists are thinking of it as not satisfying the independence requirement, thus masking substantial underlying agreement.[43]

40. See also Christensen (2007: 203; 2011: 15–17), where he appeals to the fact that an evaluation of an interlocutor's reliability might either fail to give one reason to think that one is more reliable than one's interlocutor, or, more strongly, give one positive reason to think that one is no more reliable than one's interlocutor, and claims that only in the latter case is one required to revise one's credences. Again, one might think of net resilience as a way of capturing this difference, except as a spectrum rather than a binary division.

41. Many thanks to an anonymous referee for sharpening this suggestion.

42. That said, the fact that a particular proposition seems so obvious can be of indirect relevance, since the resilience of your estimate of your own reliability will likely be especially high for especially obvious-seeming propositions. But *that* said, the resilience of your reliability estimate is certainly not a bare function of your level of credence in the proposition you are estimating your reliability with respect to.

43. This is reflected in the dialectic between Lackey (2010a: 309–310; 323–24) and Christensen. Lackey anticipates Christensen's points about your asymmetrical ability to rule out possibilities of your own unreliability via "personal information," a term which is originally hers. However, she says that, since it is

If the conciliationist says that the independence requirement is to be understood so that this reasoning does satisfy it, then the conciliationist position fully converges with my moderate, resilience-based account. Nevertheless, having already indicated why I nevertheless find talk of independence potentially misleading, there is another important point which bears stressing here. Christensen talks as if the cases in which one can downgrade one's interlocutor consistent with independence are somehow *special cases*. But once we see that they are to be understood in terms of net resilience, we see that they are not special at all; they are the norm. For *whenever* the net resilience has some positive non-zero value, *some* amount of downgrading of one's estimate of an interlocutor's reliability will be called for. And the net resilience almost always does have a positive non-zero value, since one typically possesses much better track-record data for oneself than for one's interlocutor.

This means that the independence requirement understood in this broad way will very rarely actually forbid one from downgrading one's estimate of an interlocutor's reliability in the face of disagreement, or force one to rely only on one's antecedent estimate of that reliability. Consequently, it now seems like something relatively innocuous even for a steadfast theorist to accept. After all, it is entirely consistent with the practical advice that one may remain relatively steadfast in a very wide range of cases. And it is that practical advice that many steadfasters are keen to make good on. So, it seems that once again, there is less dividing the camps than it first appears.

only in response to the occurrence of disagreement that the downgrade takes place, the reason for downgrading cannot be independent of the disagreement, whereas Christensen says that it can be. Here matters are obscured by the lack of clarity over what 'independent of the disagreement' means. Lackey is clearly right that there is a good sense in which the downgrade is not independent of the disagreement, since it takes place in response to that disagreement. But she is mistaken to think that this shows that what justifies the downgrade is your level of (justified) first-order confidence in the proposition under dispute. Rather, what she says about personal information can be understood in terms of net resilience.

7. What disagreement about disagreement II: steadfastness

This brings us to the steadfast view. Perhaps there is some theoretical commitment in this view which conflicts with the moderate view I have presented, thus dividing the camps fundamentally after all. Recall that the extreme steadfast view said that encountering disagreement, at least with anyone who you do not regard as more epistemically reliable than you, is never in itself a reason to adjust your credences. This sounds very extreme. Nevertheless, an advocate of extreme steadfastness might clarify that it is only supposed to justify steadfastness on the part of the party who had the ideal credence in light of the shared pre-existing evidence. This is sometimes called the "right reasons" view.[44]

As conciliationists like Christensen (2007: 209; 2011: 5) have pointed out, however, the very existence of the disagreement seems to be new evidence that changes the evidential position one faces compared with one's pre-disagreement situation. In the case where one's antecedent credence was the ideal credence, it is misleading evidence — but evidence nonetheless — and thus changes the ideal credence.[45] As it stands, the steadfast view is still committed to the incredible result that such misleading evidence is impossible, or that it is incapable of changing the ideal credence. For example, it's committed to denying the intuition that in **dinner check**, if you actually got the answer right, you nevertheless ought to be less confident on encountering disagreement.

To avoid this, the steadfast theorist must concede that it is your total evidence post-disagreement — including the existence of the disagreement — that determines what your post-disagreement

44. See Kelly (2005); Titelbaum (forthcoming). One might think that this dependence in verdict on whether your credence was correct pre-disagreement rather misses the point of the disagreement debate; for this criticism, see Enoch (2010: 967–68).

45. As noted in fn. 19 above, Kelly appeared to deny this in his early work, but he later concedes the point: see Kelly (2010: 136–38).

credence should be.[46] Now, however, it is unclear that the steadfast theorist is saying anything that the moderate, resilience-based view has to deny.[47] According to the resilience-based view, disagreements typically provide evidence that warrants a downgrade in credence, but they also provide evidence which warrants a downgrade in one's estimate of one's interlocutor's reliability, thus tempering the downgrade in credence. This seems entirely consistent with the claim that it is your total evidence post-disagreement that determines what your post-disagreement credence should be. Such a claim still allows for the possibility that disagreement often ought to have a significance effect on your credence.

The steadfast theorist may now claim that the difference simply lies in how significant this effect should be. In particular, the "equal weight" conciliationist view says that post-disagreement, you should split your credence between your pre-disagreement credence and that of your interlocutor. Kelly worries that this makes the evidence upon which your belief was initially based simply drop out as irrelevant, and thus amount to throwing away evidence.[48] Thus, perhaps the total evidence view is departed from after all.

As we've already seen, this equal weight extreme conciliationist view is implausible, at least as a generalization, and neglects the importance of resilience. The resilience-based view, like the total evidence view, entails that in most cases of disagreement, one should revise one's credence down somewhat, but not as far as the equal weight view would suggest. Is there nevertheless some sense in which the resilience-based view "throws away" evidence which the total evidence view does not throw away?

Perhaps it might be charged that, because the resilience-based view makes everything depend upon your antecedent credence, reliability estimates, and net resilience, it cannot give a role to what the first-order evidence regarding p actually supports.[49] This is misleading, however. The resilience-based view is supposed to give an answer to the question of how you should alter your doxastic attitudes in light of a disagreement. It is itself silent on the question of what those doxastic attitudes should be pre-disagreement, leaving that up to other doxastic norms. Nor does it say that, just by conforming to the resilience-based view, you will have an all-things-considered reasonable doxastic attitude.[50] To be sure, you will have *responded to disagreement* in the correct way, but things will have gone wrong in your doxastic states pre-disagreement so that your final attitude is nevertheless unreasonable. If you feed crazy inputs into a rational process, you may get crazy outputs: that is the fault of the inputs, not the process.[51]

Suppose that, as Kelly assumes, the doxastic norms tell you to set your pre-disagreement credence in light of your pre-disagreement

46. C.f. Kelly (2010: 141–50) and Weatherson (ms.). The Lackey (2010a, 2010b) view already considered in section 4 above is similar, though I think not identical.

47. This isn't to say that the total evidence claim is necessarily true, at least construed as a claim about rationality. (See Worsnip ms-a.) But these concerns are orthogonal to the (supposed) dispute over disagreement. The point is that if one is convinced by the total evidence claim, the resilience-based view can accommodate it.

48. See Kelly (2010: 122–25). C.f. also Weatherson (ms.).

49. To sharpen this worry, note that the resilience-based view, as I have stated it, gives a broadly symmetrical account of how the 'right' party and the 'wrong' party should respond to disagreement: So, for example, suppose that David himself has just as high a net resilience as I do: the view sketched here tells David to be just as steadfast in his view as I am in mine.

50. Kelly (2010: 127) has a further response to this move, but it is convincingly responded to in turn by Christensen (2011: 5–8).

51. Indeed, similar points apply even to an equal weight view: see Christensen (2011: 4–8). Similar issues arise with any norm that tells you what to do in light of pre-existing attitudes — for example, norms of instrumental rationality. Here, the wide-scope view (see fn. 9 above) handles things very nicely. Another possibility (Björnsson & Finlay 2010; Wedgwood ms.) is that some form of contextualism about deontic terms can be developed so that there is a more 'subjective' sense of 'should' which takes your pre-existing attitudes as fixed and tells you what to do or believe in light of them, and a more 'objective' sense of 'should' which refers to what you should do or believe given the objectively correct pre-existing attitudes.

evidence.[52] The resilience-based view then tells you how to respond to a disagreement given that pre-disagreement credence and other estimates you have. Given that, the pre-disagreement first-order evidence will in no way drop out in determining what your post-disagreement credence ought to be, since it sets the baseline from which you should be conducting revisions (as well as potentially making an indirect difference to the resilience of your estimate of your own reliability in the circumstances). The magnitude of the revision demanded by the resilience-based model does not automatically adjust such that you'll end up with the same post-disagreement credence regardless of your pre-disagreement credence.[53] So, the resilience-based view is compatible with the claim that your final credence post-disagreement should reflect your total evidence.

52. The nice thing about you, as an epistemic agent, is that you can do your best to set your pre-disagreement credence in response to *all* the pre-disagreement evidence. In this respect you are importantly different to a thermometer (c.f. White 2009; Enoch 2010), but in a way that actually makes it *easier* to line up your credence with your estimate of your own reliability than with that of a thermometer's reliability. White (2009: 241) points out that you should not always line up your credence in a thermometer's verdict with the reliability of that thermometer, since you may be aware of other evidence bearing on the temperature other than the thermometer's reading. But the same issue does not arise with respect to you, precisely because your credence is your attempt to respond to *all* the evidence of which you are aware. So, you should not find yourself wondering how to balance this credence against further evidence of which you are aware. Consequently, basing your credence on your estimate of your own reliability is at least never *knowingly* to "throw away evidence" in the way that basing a credence solely on your estimate of a thermometer's reliability might be; nor is it ever to throw away evidence that you had previously responded to.

53. Once more, the same goes for the equal weight view. It might be tempting to think that the equal weight view tells you to move to credence 0.5, so it doesn't matter what your pre-disagreement credence was: you'll always end up responding to disagreement the same way. But that's not what the equal weight view says. The equal weight view tells you to split the difference between your credence and that of your interlocutor. It only tells you to move to credence 0.5 in the special case where you and your interlocutor have credences that are equal distance from 0.5 on different sides. Holding fixed your interlocutor's credence, your pre-disagreement credence certainly *will* make a difference to what your post-disagreement credence should be on the equal weight view.

A steadfast theorist might still object, however, that there is something important separating the resilience-based view and the steadfast view. According to the resilience-based view, one needs to have a high net resilience to resist revising one's credence in response to a disagreement. A steadfast theorist might now charge that, although net resilience can explain why one should not conciliate in a significant range of cases, there are other cases where one should not conciliate: cases where one's first-order evidence for one's original credence is simply overwhelmingly strong. The steadfast theorist might hold that one should remain steadfast in such cases even if one's net resilience is not high.[54]

One might think that it would be easy to construct cases where one's first-order evidence for one's original credence is very strong, but one's net resilience is low. Actually, though, it is not as easy as it seems. For in general, if one's first-order evidence for one's original credence is very strong, and one recognizes this, one ought to have a relatively high net resilience. Here is why. Remember that the reliability at issue when one makes reliability estimates is reliability *in the circumstances*. If we are discussing baseball, I am interested in your reliability on matters concerning baseball, not the overall reliability of your beliefs. Now, when it comes to some matter on which one has overwhelmingly strong evidence, one generally has good reason to be extremely resilient about one's own reliability. So, for example, consider elementary mathematical calculation, where basic sums are overwhelmingly obvious. These are amongst the cases where it is easiest to know that one is reliable: where it is not only the case that one is very reliable, but where one is extremely confident in one's own reliability.

Now, one may also have a relatively resilient estimate of an interlocutor's reliability on such matters. But it will not be as resilient as one's estimate of one's own reliability — for the sorts of reasons that Christensen drew our attention to: one cannot eliminate the

54. I am very grateful to an anonymous referee for pressing this objection.

possibility that one's interlocutor has suffered a temporary fit of madness, or that he is on drugs, or something similar, in the way that one can eliminate these possibilities for oneself. So, given that one's estimate of one's own reliability is *so* resilient, the net resilience is still significantly above zero. Given these facts, the resilience-based view does predict the steadfast theorist's result that in most cases where one has overwhelmingly strong evidence, one may remain steadfast.

Of course, the steadfast theorist could now try just stipulating that in some particular case, one has overwhelmingly strong first-order evidence but the resiliences are the same: one is no better able to rule out the possibility that one have oneself suffered a fit of madness than one can rule that possibility out for one's friend. I do not claim that this is impossible; merely very atypical. But when this very atypical case is explicitly stipulated, it does not seem intuitively attractive to me that one should remain steadfast. Again, it's not that the first-order evidence gets thrown away here; rather, it's just that in such a case, the disagreement really is quite strong (even if ultimately misleading) countervailing evidence that you yourself may be the one who has suffered the fit of madness. The strength of this countervailing evidence cannot just be stipulated away by the steadfast theorist. And it calls for a significant reduction in credence, by the very lights of the total evidence view.[55] So, the unusual case in which one has strong first-order evidence but a low net resilience is not one where one should remain steadfast, even by the lights of the total evidence view.[56]

55. See Christensen (2007: 200) for another very compelling case where one's pre-disagreement evidence overwhelmingly supports a proposition, but one should conciliate significantly upon encountering disagreement. Again, this case is plausibly read as exemplifying the unusual pattern where one has a low net resilience despite having very strong first-order evidence for one's view.

56. As we saw in section 4, a case where one has a high net resilience just is one where one should downgrade one's estimate of one's interlocutor's reliability. So, the fact that one may not remain steadfast despite a low net resilience means that one cannot have a case where one should remain steadfast without downgrading one's estimate of one's interlocutor's reliability, chalking it up merely to an uncharacteristic error on one's interlocutor's part.

So, while the steadfast theorist *could* distance herself from the resilience-based view by claiming that one may remain steadfast even when the net resilience is low, I do not think that she *should* do this. The resilience-based view can get her the practical result she wants that in a very wide range of cases, one is permitted to remain relatively steadfast; including the vast majority of cases where one's first-order evidence supports one's view overwhelmingly. And the cases where the resilience-based view does not support steadfastness are ones where the disagreement really is very strong countervailing evidence against one's first-order evidence. And that is exactly the result which a fan of the total evidence view should want.

Both the total evidence view and the resilience-based view, then, can think of you as starting with a pre-disagreement credence based on your pre-disagreement evidence and then responding to the new evidence you gain from the disagreement. And, as we've seen, Christensen's moderation of the conciliationist view collapses, on its most charitable interpretation, into the resilience-based view. Surprisingly, then, we find nothing that identifiably separates the moderate version of the steadfast view from the moderate version of its conciliationist rival, and nothing that prevents either party from embracing the moderate, resilience-based view. Even if it turns out that there are some differences to be found, though, the moderate account does well explaining and accommodating the intuitions behind both sides. As things stand, it seems to me a very promising candidate to resolve the (seeming) disagreement about disagreement.[57]

57. In writing this paper, I have benefitted from conversations with David Black, David Christensen, Steve Darwall, Keith DeRose, Georgi Gardiner, Alvin Goldman, John Pittard, Sander Verhaegh, and Steve Yablo. I am also very grateful to two extraordinarily helpful anonymous referees at *Philosophers' Imprint* for written comments. Thanks also to participants at a conference hosted by the University of Miami, where an earlier version of the paper was presented, and to Micah Dugas for serving as the commentator there.

References

Adams, E.W. (1965). "The Logic of Conditionals," *Inquiry*, 8: 166–197.

Björnsson, G. & Finlay, S. (2010). "Metaethical Contextualism Defended," *Ethics*, 121/1: 7–36.

Broome, J. (1999). "Normative requirements," *Ratio*, 12: 398–419.

———. (2013). *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.

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

———. (2011). "Disagreement, Question-Begging and Epistemic Self-Criticism," *Philosophers' Imprint*, 11/6.

Dancy, J. (2000). *Practical Reality*. Oxford: Oxford University Press.

Elga, A. (2007). "Reflection and Disagreement," *Noûs*, 41/3: 478–502.

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

Feldman, R. (2006). "Epistemological Puzzles about Disagreement," in Hetherington (ed.), *Epistemology Futures*. Oxford: Oxford University Press.

Feldman, R. & Warfield, T.A. (eds.) (2010). *Disagreement*. Oxford: Oxford University Press.

Joyce, J.M. (2005). "How Probabilities Reflect Evidence," *Philosophical Perspectives*, 19: 153–178.

Kaufmann, S. (2004). "Conditioning Against the Grain," *Journal of Philosophical Logic*, 33: 583–606.

Kelly, T. (2005). "The Epistemic Significance of Disagreement," *Oxford Studies in Epistemology*, 1: 167–196.

———. (2010). "Peer Disagreement and Higher-Order Evidence," in Feldman & Warfield (2010).

———. (2013). "Disagreement and the Burdens of Judgment," in Christensen & Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

King, N.L. (2012). "Disagreement: What's the Problem? *Or* A Good Peer is Hard to Find," *Philosophy & Phenomenological Research*, 85/2: 249–272.

Lackey, J. (2010a). "A Justificationist View of Disagreement's Epistemic Significance," in Haddock, Millar & Pritchard (eds.), *Social Epistemology*. Oxford: Oxford University Press.

———. (2010b). "What Should We Do When We Disagree?" *Oxford Studies in Epistemology*, 3: 274–293.

Lam, B. (2013). "Calibrated probabilities and the epistemology of disagreement," *Synthese*, 190/6: 1079–1098.

McGee, V. (2000). "To tell the truth about conditionals," *Analysis*, 60/1: 107–111.

Skyrms, B. (1980). *Causal Necessity*. New Haven: Yale University Press.

Titelbaum, M. (forthcoming). "Rationality's Fixed Point (Or: In Defense of Right Reason)," *Oxford Studies in Epistemology*, 5.

Van Inwagen, P. (2010). "We're Right. They're Wrong," in Feldman & Warfield (2010).

Weatherson, B. (ms.). "Do Judgments Screen Evidence?," draft manuscript, University of Michigan.

Wedgwood, R. (ms.). "Objective and Subjective 'Ought'," draft manuscript, University of Southern California.

White, R. (2009). "On Treating Oneself and Others as Thermometers," *Episteme*, 6/3: 233–250.

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

———. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.

Worsnip, A. (ms-a). "The Irrelevance of Evidence (to Rationality)," draft manuscript, Yale University.

———. (ms-b). *Rationality's Demands on Belief*. Draft dissertation, Yale University.