Canadian Journal of Philosophy (2025), 55: 5-6, 452–468
doi:10.1017/can.2026.10070

ARTICLE

The Rational Response to Excessive Confidence[1]

Thomas Mulligan

Georgetown University, Washington, USA
Email: thomas.mulligan@georgetown.edu

Abstract

Conciliationism holds that it is rational to modify one's beliefs in the face of disagreement. But extant conciliatory norms yield incorrect results in cases involving *excessive confidence*—cases in which one's interlocutors are sure, or almost sure, that their opinions are correct. After explaining the problem of excessive confidence, I show that a Bayesian approach to Conciliationism handles the problem elegantly and effectively. Further, it has desirable—indeed essential—features, including the ability to (i) contend with multiple interlocutors, (ii) deal with gradations in competence, (iii) incorporate ubiquitous interdependence, and (iv) account for critical contextual features of cases of disagreement.

Keywords: disagreement; Conciliationism; excessive confidence; Bayesianism

Ann, a philosopher, isn't sure if God exists. One day, Ann meets another philosopher whom she regards as equally smart and equally familiar with the evidence and arguments that bear on the existence of God. Ann asks the other philosopher, "Do you believe in God?"

"No," the other philosopher says.

"How confident are you in your judgment?"

"Utterly."

"You don't think that there's *any* possibility that God exists?"

"That's right," the other philosopher says.

Nancy, a schoolteacher, is confident that the Republican will lose the next election. One day, Nancy meets another schoolteacher whom she regards as equally smart and equally familiar with political matters.

"I think that the Republican has a 20% chance of winning–what do you think?" Nancy asks.

"100%."

[1]I thank Jesse Hill, Mike West, and the audience at the 2021 Joint Session of the Aristotelian Society & Mind Association for helpful comments on the matters discussed herein. The submitted (and accepted) version of this article was prepared without the use of artificial intelligence. AI (Claude Opus 4.7 and ChatGPT 5.5) was used in making final revisions to the manuscript. This help included double-checking calculations, identifying typos, searching for relevant empirical literature on overconfidence (§1), and writing R code to regenerate the two figures in Section 4 at higher resolution. These AIs also made helpful suggestions about incorporating referee comments into the final version of the manuscript, about n. 18, and about treating λ as a function of x (Section 4). All final content is the author's responsibility, alone.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Canadian Journal of Philosophy, Inc. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

"You don't think that there's *any* possibility that she'll lose?"
"None whatsoever," the other schoolteacher says.

These are two cases of disagreement with an epistemic peer.[2] Otherwise unremarkable, they are distinguished from the myriad cases in the epistemology of disagreement literature by the extreme confidences that the interlocutors have in their beliefs.[3] In the first case, the other philosopher has perfect confidence that *God exists* is false; in the second case, the other schoolteacher has perfect confidence that *the Republican will win the next election* is true.

My purpose in this essay is to argue that cases of *excessive confidence* like these imperil conciliatory theories of disagreement, as currently construed. These theories call for a person to modify her confidence in a proposition when faced with disagreement from an epistemic peer.[4] This is usually sensible advice, but it is too quick when an epistemic peer is sure, or almost sure, that his opinion is correct.

Excessive confidences must be handled differently by our conciliatory norms. The rational response in these cases is—roughly speaking—to remain "steadfast" in one's belief *to some degree*. This is because excessive confidences are irrational, and we should take an interlocutor less seriously—perhaps not seriously at all—when he expresses one. Since the most popular conciliatory norms (e.g., those discussed at the beginning of Section 1) treat extreme confidences (say, 0.9999) and temperate confidences (0.7) the same, none is fully general, nor fully correct.

I proceed as follows: In Section 1, I describe the conciliatory position and present three conciliatory belief revision norms. I also say a few words about why excessive confidence is irrational. In Section 2, I show that cases of excessive confidence (i) are not mere mathematical pathologies, (ii) are orthogonal to the "Independence" debate, and (iii) cannot be easily dismissed by claiming that anyone who expresses an excessive confidence is not, *ipso facto*, an epistemic peer. In Section 3, I present two test cases that any approach to excessive confidence should be able to contend with. In Section 4, I describe how our conciliatory norms may be modified to deal with the problem of excessive confidence. I argue that a Bayesian approach to the epistemology of disagreement handles the problem elegantly and effectively. I discuss some ramifications of the preceding analysis for the general disagreement debate in Section 5. I conclude in Section 6.

One preliminary note. To this point, I've talked about cases of disagreement with an epistemic peer. But as I have argued before (Mulligan, 2021), the peerhood case is an artificial one. It seems that many, maybe most, of our epistemic interactions involve people who are at least a little less, or a little more, "competent" or "accurate" (etc.) This essay's argument is perfectly general. If we ought, normally, to defer to an epistemic interlocutor who is our superior to a degree δ, then—all else equal—when that interlocutor is excessively confident, that degree of deference should drop to $\gamma < \delta$. This raises the possibility that we might show more deference to a non-excessively confident epistemic peer than to an excessively confident epistemic superior.

[2]Kelly gives a classic definition of *epistemic peerhood*: "Let us say that two people are **epistemic peers** with respect to some question if and only if they satisfy the following two conditions: (i) they are equals with respect to their familiarity with the evidence and arguments which bear on that question, and (ii) they are equals with respect to general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias." (2005: 174–75). Now, other definitions of peerhood (e.g., Elga's [2007]) are "conditional" (on the disagreement at hand). And on such definitions, it might be argued that an "excessively confident peer" is something like a contradiction-in-terms. But that does nothing to eliminate the problem raised in this article; it merely relocates it. As I'll discuss in Section 2, disagreement norms must be general enough to contend with people who are not strictly our peers, and thus we must grapple with the question of how much weight to afford excessively confident epistemic inferiors and superiors.

[3]I shall use "confidence" to refer to the (subjective) probability that a person assigns to a proposition. This is common usage in the literature, although sometimes "credence" is used instead, and, in some of those cases, "confidence" plays a different role, as a higher-order concept.

[4]See, for example, Christensen, 2007, Elga, 2007, and Feldman, 2007.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

1. Conciliationism and Excessive Confidences

A *disagreement norm* is a function $f : [0,1] \times [0,1] \to [0,1]$.[5] Given her own confidence c in a proposition and her interlocutor's expressed confidence x, a person may compute a unique value $f(c,x)$ to adopt as her new, post-disagreement confidence in the proposition.

A number of disagreement norms have been proposed in the conciliatory literature. For example, there is the "Equal Weight View" (or what Jehle and Fitelson [2009] call "Straight Averaging"):

$$f_1(c,x) = \frac{c+x}{2}. \tag{1}$$

There is Easwaran et al.'s (2016) "Upco" ("Updating on the Credences of Others"):[6]

$$f_2(c,x) = \frac{cx}{cx + (1-c)(1-x)}. \tag{2}$$

There is my Bayesian preference (Mulligan, 2021), which was derived by Genest and Schervish (1985):

$$f_3(c,x) = c + \lambda(x-\mu), \tag{3}$$

where μ and λ are parameters to be discussed in Section 4.

Although each norm will typically produce a different post-disagreement confidence, they share a feature relevant for this article: A single function applies to the entire domain. That is, disagreement norms are not piecewise functions.

Why should we distinguish excessive confidences, such as $x = 0$ and $x = 1$, from other values in the interval? And why should we not conciliate as usual when we disagree with someone who expresses an excessive confidence? There are four reasons.

First, excessive confidence is evidence of cognitive bias on the part of the person expressing it. The existence of God is, to put it mildly, a difficult and complicated matter. Our brightest minds have struggled with it for millennia, without resolution (some would say without progress). Even the most devout among us ought to profess to something less than certainty in his belief in God.

Similarly, predicting elections is a tricky business. The most sophisticated models, fed with ample data, regularly get things wrong. Thus, when one hears an interlocutor say that he is *absolutely sure* that God exists (or that He does not), or that the Republican will lose the election (or win it), we learn that he is badly afflicted by wishful thinking.

Second, excessive confidence does not make sense from a Bayesian point of view. To say, for example, that a person has a confidence of 1 in some proposition is to say that there is *no* future evidence that could shake him from that belief. But that should not be the case in the strongest real-world circumstances. So it should not ever be the case.[7]

The strongest real-world circumstances are mathematical proofs. If there's a case in which we can be confident that a proposition (theorem) is true, it is when it has been demonstrated rigorously. But even when we have such a proof, that does not justify perfect confidence in the proposition. This is because we might one day discover that the proof contains an error.

[5]The disagreement literature typically considers the special case in which one is entertaining disagreement with a single epistemic peer. For the general case of disagreement with n peers, $f : [0,1]^{n+1} \to [0,1]$.

[6]*N.B.* Upco is only partly defined on the closed square; for example, it is undefined when $x = 0$ and $c = 1$.

[7]"So leave a little probability for the moon to be made of green cheese; it can be as small as 1 in a million, but have it there since otherwise an army of astronauts returning with samples of the said cheese will leave you unmoved" (Lindley, 1971: 105–6).

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

For example, it was "proven" (by Euler, no less) that for all polyhedra, the number of vertices of a polyhedron, plus the number of its faces, minus the number of its edges, equals 2. This result was widely accepted by mathematicians for years. If you had asked them, some probably would have professed perfect confidence in it because it had been demonstrated *via* the method of rigorous proof.

The problem, however, is that the proof was erroneous (this formula, the so-called "Euler characteristic," only holds for *convex* polyhedra).[8] Although an eighteenth-century mathematician might not have been able to point to *the* flaw in Euler's proof, he could have, and should have, known that the existence of a flaw *was possible.* And he should not have had perfect certainty in the theorem as a result.

Third, there are skeptical considerations that speak against excessive confidence. In the *Meditations*, for example, Descartes raises the possibility that we are being manipulated by an evil demon to believe (what strike us as) obvious truths, like $2 + 3 = 5$.

Now Descartes, of course, believes that knowledge of this proposition can be recovered through the *cogito*. But even if you buy that reasoning, or other anti-skeptical arguments, you might be wrong. And even if the chance that you are wrong is tiny, that is sufficient to make excessive confidence untenable.

Finally, fourth, the problem of excessive confidence is not merely theoretical. It is a ubiquitous feature of human judgment which will taint many instances of real-world disagreement. In the words of one social psychologist, "no problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence" (Plous 1993: 217).

For example, one classic study (Fischhoff et al., 1977) found that experimental subjects expressing odds of 1,000:1 that their answers were correct (∼99.9% confidence) were nevertheless wrong 12–19% of the time, across three experiments. The calibrated odds for their actual accuracy were about 6:1.[9] Because excessive confidences are frequently miscalibrated, it is a mistake to reflexively let them influence our beliefs more than temperate confidences do. In sum, any theory of disagreement that handles excessive confidences badly will handle a significant fraction of real-world cases of disagreement badly.

2. What the Problem Is Not

Three immediate and easy resolutions to the problem spring to mind. None works.

It might be thought, first, that this is just a mathematical pathology—a result of drawing confidences from the closed interval $[0,1]$ rather than the open interval $(0,1)$.

However, very low (but $\neq 0$) and very high (but $\neq 1$) confidences pose the same problem. There is nothing relevantly different about an interlocutor reporting a confidence of 0.0001 in *the Democrat will win the election* and a confidence of 0. The former is to say that the probability of a Democratic victory is akin to the probability that you'll be struck by lightning,[10] which is ridiculous.

Indeed, in some contexts even more modest confidences—say, 0.9—can qualify as excessive. If the proposition in question is *the stock market will go up tomorrow*—a notoriously hard thing to predict—then a confidence of 0.9 may suggest irrationality.[11] As will be discussed, context matters critically.

[8]For some history, see Francese and Richeson (2007).

[9]See Moore and Healy (2008) for a summary of the overconfidence literature.

[10]See https://www.weather.gov/safety/lightning-odds, retrieved 20 September 2024.

[11]Of course, the case can be manipulated such that this degree of confidence in *the stock market will go up tomorrow* may be warranted. For example, if you sit on the board of a central bank, and the board just met and decided to cut interest rates (and will announce this in the morning), then such a high confidence may not be excessive. But that is not the case at hand.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

Second, it might be thought that the problem raised in this essay is subsumed by the "Independence" debate in the epistemology of disagreement. In fact, it is independent, as it were, of that debate.

The Independence Principle holds that one must not use the fact of disagreement to discount an interlocutor's opinion. As David Christensen puts it, Independence "prevent[s] blatantly question-begging dismissals of the evidence provided by the disagreement of others. It attempts to capture what would be wrong with a P-believer saying, *e.g.*, 'Well, so-and-so disagrees with me about P. But since P is true, she's wrong about P. So however reliable she may generally be, I needn't take her disagreement about P as any reason at all to question my belief.'" (2011: 2).

In the cases under consideration here, however, it is not disagreement *per se*, or the mere fact that an interlocutor has a confidence different from our own, that leads us to discount his opinion. It is that there is something peculiar about excessive confidence; namely, it is strongly suggestive of irrationality. When we (with a confidence of, say, 0.4 in some proposition) ignore an interlocutor who expresses an extreme confidence (0.9999), we do not do so because he disagrees with us. (Observe that we would not dismiss him if he had a confidence of 0.9. We would conciliate as usual.) The disagreement is besides the point. It is the *confidence*, rather, and what it implies, that generates the dismissal.

Third, it is tempting to say that these cases of disagreement involving excessive confidence are moot since there is no disagreement with an epistemic peer in the first place, since no epistemic peer can be excessively confident. After all, I have argued that interlocutors who express an excessive confidence are not equally thoughtful, nor equally free from bias. And on a plausible definition of epistemic peerhood (*cf.* n. 2), that disqualifies them from peerhood status. Alas, this does not solve the problem.

For one thing, I think some will intuit the opposite: that people whom we regarded as epistemic peers before disagreement remain peers after they express an excessive confidence. What is unique about the situations we are considering is that it is the *confidence itself*—the value of x—that is problematic. This is categorically different from typical facts that speak to peerhood status, all of which involve the character of an interlocutor, the way he expresses confidence, and so on. For example, a significant difference in education or in intellectual capability may disqualify someone from peerhood status. So may the manner in which x is asserted: If our (otherwise-qualified) interlocutor says "x" with slurred speech, glassy eyes, and breath redolent of alcohol, that may mean that he is not an epistemic peer after all (because he is drunk). But if we imagine a real-world case akin to the first vignette—we meet another philosopher whom we believe understands the relevant issue as well as we do, has published on it equally, seems about as smart as we are, has never expressed bias as best we can tell, and so on–well, some may retain the intuition that this person is a peer.[12]

That there may be something to this intuition may be clearer with a case of epistemic superiority. Suppose we meet an accomplished mathematician, a number theorist, and we ask her, "What is your confidence in *Fermat's Last Theorem is true*?" She replies that she is perfectly confident in this proposition; after all, the theorem was proven in 1995 by Andrew Wiles (with help from Richard Taylor).

In this case, one cannot plausibly argue that our interlocutor is not at least our epistemic peer. Indeed, it seems clear that she is our epistemic superior. Yet there is still something wrong with her extreme confidence. After all, the two relevant papers (Taylor & Wiles, 1995; Wiles, 1995) are 129 pages long and contain methods only a handful of mathematicians understand. Even though he

[12]Is there anything analogous to the problem of excessive confidence, discussed in this essay? One possibility is a problem of *excessive precision*. Suppose that, in response to the proposition *the Mets will win tomorrow*, your interlocutor gives a confidence of 0.781652. Here, too, there is something about the confidence itself that implicates an epistemic problem in the person uttering it. It is irrational to have such a precise judgment in a context like this. I thank an anonymous referee for suggesting this possibility to me.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

was confident it was sound, Wiles's original proof contained a fatal error (caught by referees at the first journal he approached). It is possible that another, subtle error snuck through. Moreover, Wiles's proof relies on many results from other mathematicians, which in turn rely on others' results, and there could be an error in one of those papers. If any of that is possible, then perfect confidence in the proof is irrational.

The point, then, is that the irrationality of excessive confidence is different than the irrationality that implicates peerhood. To fully explore this matter would require a conceptual analysis of epistemic peerhood (on that see, e.g., Gelfert, 2011; Lougheed, 2020), but that is unnecessary, since even if one *does* think that excessive confidence imperils peerhood status, that does not eliminate the problem I have raised.

The reason is that, as alluded to in the introduction, it is rarely the case that a person is a peer *simpliciter*. Rather, our interlocutors tend to be a little smarter or a little less smart, a little more knowledgeable along this line but less knowledgeable along that one, and so on. But no matter what our interlocutor is—epistemic inferior, peer, or superior—we have reason to be skeptical of his opinion if it involves excessive confidence. So, merely disqualifying him as *a peer* for expressing excessive confidence is not enough, since we would still end up modifying our belief too much (albeit using a different norm) as a result of disagreement.

Indeed, if we consider the case of Fermat's Last Theorem, above, it seems that we should give some deference to our interlocutor's opinion when she expresses perfect confidence in *Fermat's Last Theorem is true.* We should increase our confidence in this proposition. But we should not increase our confidence as much as we would have, had she given a high but not excessive confidence (say, $x = 0.98$).

3. Two Test Cases

I will now consider two objections to the general approach to dealing with excessive confidence which I have outlined.[13] I shall not resolve them here; rather, we shall use them, in the following section, as test cases as we consider how to rigorously treat the problem of excessive confidence.

Objection 1. Suppose confidences above 0.9 are excessive when it comes to proposition *X*, but your confidence, 0.85, is not. You meet two people, one of whom has a confidence of 0.2 in *X* and the other of whom has a confidence of 0.91. Although excessively confident, the second person's opinion cannot be wildly off (being so similar to your own view, that would mean that your own confidence is wildly off—which is epistemically untenable). So the second person's opinion should count for *something*.

Objection 2. Suppose confidences above 0.9 are excessive when it comes to proposition *Y*, but your confidence, 0.2, is not. You meet five people, all of whom have confidences of 0.91 in *Y*. Even if these five people are excessively confident, their consensus surely counts for *something*. Indeed, it seems epistemically important.[14]

4. Discussion

It is irrational to be excessively confident in one's beliefs. Thus, when we encounter an excessively confident interlocutor within the context of disagreement, we gain insight into his epistemic status. That insight, in turn, justifies an extra degree of skepticism.

[13]I thank an anonymous referee for posing these cases to me.

[14]That said, there are cases in which consensus—even overwhelming consensus—is *not* epistemically important; most notably, cases in which the judgments of the group members are interdependent. In the special case of perfect dependence, consensus counts for nothing, epistemically. Dependence will be discussed in more detail in Section 4.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

How, then, should Conciliationists modify our theories to deal with the phenomenon of excessive confidence? A straightforward approach is to simply ignore those confidences that you regard as excessive. Suppose, for example, you think that Conciliationism's Equal Weight View (equation (1)) is correct in its broad strokes. That is, except for cases involving excessive confidence, we ought to "meet in the middle" with our interlocutor. Then we could define our disagreement norm piecewise, as:

$$f_1'(c,x) = \begin{cases} c & \text{if } x < \varepsilon, \\ c & \text{if } x > 1-\varepsilon, \text{and} \\ \dfrac{c+x}{2} & \text{otherwise,} \end{cases} \tag{4}$$

where $\varepsilon \in (0,0.5)$ defines excessive confidence in the given context. For example, suppose that we take $\varepsilon = 0.02$. Then, if our interlocutor says he is $>98\%$ sure that the proposition is true, we ignore his opinion and remain steadfast. If he says he is $<2\%$ sure we ignore his opinion and remain steadfast. For any other $x \in [0.02,0.98]$, we conciliate as usual.[15]

One problem with this norm is that it blunders into the two objections outlined in the last section. In those cases, $\varepsilon := 0.1$. Considering the first objection: This norm fails to grant the (slightly excessive) interlocutor (with confidence 0.91) any epistemic influence at all. Whether he is there or not is irrelevant—we end up with a confidence of 0.53. And in the case of the second objection (five interlocutors with confidences of 0.91 in Y), there is no confidence change at all—which, as mentioned, seems wrong.

Further, this norm is too crude. It stands to reason that one should be "more steadfast" in the face of a confidence like $x = 1$ than in the face of $x = 0.99$, even if both are above the threshold $1-\varepsilon$. Equation (4) does not capture that dynamic.

More generally, defenders of approaches like the Equal Weight View should aspire to a norm that captures gradations in epistemic peerhood. If the norm can do that, then degrees of excessiveness (and, thereby, degrees of irrationality) can be captured as well.

Because I think that ultimately this is not the most promising route for the Conciliationist to take, I shall not explore the development of such a norm in detail. However, an example will illustrate what I have in mind.

First, define a scaling function

$$K(x) := \left(-4x^2 + 4x\right)^{\varepsilon}, \tag{5}$$

where $\varepsilon \in (0,\infty)$ is interpreted similarly to above: The greater ε be, the larger the range of confidences we wish to regard as excessive in the given context (see Figure 1). ($K(x)$ is the scaled Bernoulli variance raised to ε.)

Now modify equation (1) as follows:

$$f_1''(c,x) = c + K(x)\frac{x-c}{2}. \tag{6}$$

This is the Equal Weight View, but modified such that excessive confidences do not lead to as much belief revision as more temperate confidences do. What counts as "excessive" and "temperate" is determined by the ε parameter. In the most extreme cases, in which $x = 0$ and $x = 1$, there is no belief revision at all; we remain steadfast.

[15]It is not necessary that the same value for ε be used at the top of the interval and the bottom. There could be a case, for example, in which (i) confidences greater than 98% are excessive and (ii) confidences less than 10% are excessive.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

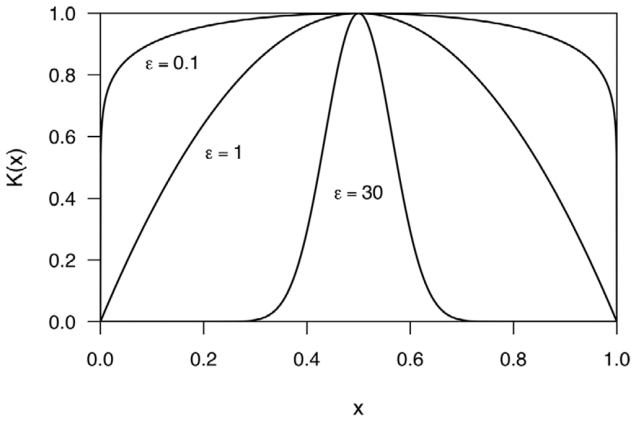


Figure 1. Scaling function K

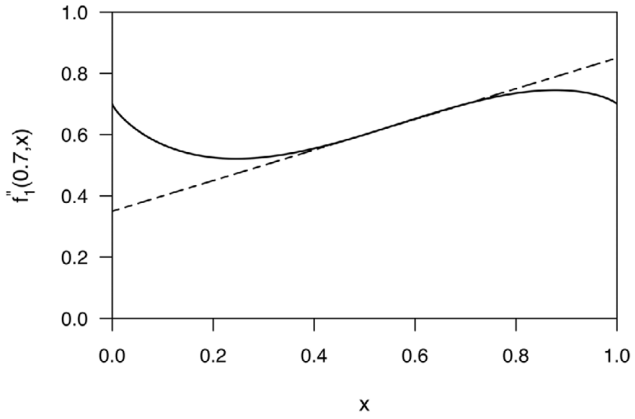


Figure 2. Modified Equal Weight View ($c = 0.7$; $\varepsilon = 0.8$)

Figure 2 describes the norm for $c = 0.7$ and $\varepsilon = 0.8$. The dotted segment represents the classic Equal Weight View (i.e., equation (1)):

The modified Equal Weight View f_1'' approximates the classic Equal Weight View f_1 well for most of its domain. It diverges, however, at the extremes—which is the desired behavior. For example, $f_1''(0.7,0.7) = f_1(0.7,0.7) = 0.7$; $f_1''(0.7,0.4) \approx f_1(0.7,0.4) = 0.55$; $f_1''(0.7,0.99) = 0.71$ (whereas $f_1(0.7,0.99) = 0.85$); $f_1''(0.7,0.01) = 0.67$ (whereas $f_1(0.7,0.01) = 0.36$).[16]

How does the modified Equal Weight View deal with the two objections of §3? Let us begin with Objection 2. Since we are dealing with multiple interlocutors, we cannot directly apply equation (6). But since they all share the same confidence (x_1=x_2=x_3=x_4=x_5=0.91), there is a natural modification:

$$f_1'''(c,x_1,x_2,x_3,x_4,x_5) = c + K(x_1)\frac{5(x_1 - c)}{6}. \tag{7}$$

[16]The scaling function K is defined as it is for simplicity's sake. Perhaps it can be changed to incorporate further behavior (e.g., asymmetry about $x = 0.5$) and to approximate the classic Equal Weight View to arbitrary accuracy.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

Choosing $\varepsilon = 0.2$, this yields a post-disagreement confidence of 0.67, which seems reasonable (i.e., going from 0.2 to 0.67 when faced with five highly confident peers whose confidence is, nevertheless, irrationally high.)

Indeed, this looks like a good result when compared to the simplest norm for dealing with multiple interlocutors (and the natural generalization of the classic Equal Weight View)—linear averaging. That yields a post-disagreement confidence of 0.79. For the reasons given in this article, that value is too high because of the fact, not contemplated by linear averaging, that these peers' confidences are excessive. Equation (7) gets you the haircut you are looking for.

Now consider Objection 1. In this case, our interlocutors do not share the same confidence (unlike in Objection 2), and a tempting modification of equation (6) is:[17]

$$f''''_1(c, x_1, x_2) = c + \frac{x_1 - c}{3} + K(x_2)\frac{x_2 - c}{3}, \tag{8}$$

where x_2 is the excessively confident interlocutor (i.e., the one with a confidence of 0.91).

Again choosing $\varepsilon = 0.2$, we have a post-disagreement confidence of 0.65, which happens to be the same (with rounding) as the result of linear averaging. So this, too, seems sensible.

But now there is a conceptual problem. Equation (8) creates an asymmetry between yourself and x_1 which is not justified. To illustrate, suppose we are in a context in which almost no weight should be given to the opinion of a highly confident peer. This may be expressed by setting $\varepsilon = 7$, for instance (which essentially eliminates the third term on the right-hand side of equation (8)).

Then equation (8) yields a post-disagreement confidence of 0.63 (whereas the theoretically appropriate value is 0.53, splitting the difference between yourself and your (non-excessively confident) interlocutor, who has a confidence of $x_1 = 0.2$). The problem is the 3 in the denominator in the second term on the right-hand side of equation (8). All of the influence which x_2 would have exerted, had he not been excessively confident, is in effect replaced by your own opinion, when it should be split up equally between you and x_1. So in fact, despite being *explicitly* ignored, x_2 continues to exert latent epistemic influence, reducing the weight you put on x_1. That is uncalled for.[18]

Taking a step back, one might object that all these modifications lack the elegance of the classic Equal Weight View, so discussed in the epistemological literature. But to my mind, there is no reason to expect that a simple, parameter-free norm like the classic Equal Weight View will capture the richness of real-world disagreement. Conciliationists attracted to generalizing their norms in this way should not be cowed by this critique.

In any event, I come down on the side of the Bayesians, in part for how they can handle these complexities. Consider again equation (3):[19]

$$f_3(c, x) = c + \lambda(x - \mu).$$

[17]In these cases, the problem of non-commutativity (see, e.g., Gardiner, 2014; Kuan, 2021) might arise. We elide this by assuming that the two confidences are solicited at the same time, which is compatible with the objection as written.

[18]Now, this defect is not inevitable for Equal Weight View—style approaches. For instance, consider this normalized modification of equation (6): $f''''_1(c, x_1, x_2) = c + \frac{(x_1 - c) + K(x_2)(x_2 - c)}{2 + K(x_2)}$, which avoids the latent epistemic influence problem. I do not pursue this family of approaches, since the Bayesian approach discussed below provides a more general and, in my view, principled way to deal with these cases.

[19]In what way is equation (3) "Bayesian"? After all, Bayes's Law provides the ideal formalization for conditionalizing on others' opinions (which are treated as just another form of data relevant to our beliefs). The problem is that applying Bayes's Law in this way requires a full, joint distribution of our interlocutors' possible reported beliefs and the underlying proposition. That means assessing the calibration and bias of our interlocutors, adjudicating issues of dependence, considering whether an "*X*%" report from one interlocutor means the same as "*X*%" from another, and so forth. It cannot practically be done. Genest and Schervish's (1985) core insight is that we need not specify all that. Instead, if we specify the marginal expectation μ (or the expectations μ_i in the case of multiple interlocutors) and then impose Bayesian-compatible consistency conditions—which we *can*, in practice, do—then the posterior updating function takes the form of equation (3).

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

The parameter $\mu \in (0,1)$ represents the confidence that a person expects her interlocutor to report. (And thus, as is the Bayesians' wont, $f_3(c,x) = c$ when our interlocutor reports what we expect her to, $x = \mu$. For in such a case, our interlocutor's belief x was baked into our prior.)

The parameter λ has several interpretations, but can be thought of as a measure of our interlocutor's reliability. When $\lambda = 0$, we assess our interlocutor's judgment as wholly unconnected to the true state of the world. Thus, we do not allow that judgment to affect our belief. When λ is negative, we assess that her judgment varies inversely with the true state of the world. When λ is positive, we conciliate in our interlocutor's direction. And the more sensitive our interlocutor's judgment is to the true state of the world, the larger λ be.[20]

Thus, the Bayesian may simply choose a lower λ than she otherwise would when her interlocutor reports an excessive confidence. If she thinks it appropriate to remain steadfast in the face of excessive confidence, then she sets $\lambda = 0$. If she thinks some conciliation is appropriate, she chooses a positive λ. In either case, equation (3) works.

There are two complications. First, we would like to select a value for λ before x is revealed to us; that way, Independence-type worries (Section 2) do not arise. But now this cannot be done; we cannot finalize λ until we hear our interlocutor's opinion.

Second, West and Crosse (1992) suggest a promising approach to the selection of λ. Observe from equation (3) that we can work backwards, computing λ by deciding on an appropriate posterior probability, $f_3(c,x)$, for a given value of x. This single assessment fixes the putatively proper value of λ no matter what our interlocutor actually says.

West and Crosse suggest (p. 288) using the extreme cases of $x = 1$ or $x = 0$, or $x = 1 - \varepsilon$ or $x = \varepsilon$ (where ε is small and positive). Put in the lingo of this essay, they suggest that we choose λ by considering what we ought to believe, were we to hear an excessive confidence.

This approach will not work. Suppose that, following the argument of this essay, we judge that if we hear one of these excessive confidences, then our opinion should be unchanged. Notice, then, that we have

$$c = c + \lambda(x - \mu) \Rightarrow \lambda = 0 \vee x = \mu \Rightarrow f_3(c,x) = c.$$

In other words, by following this advice for selecting λ, we are driven to Steadfastness *no matter what our interlocutor actually reports*. But that is not right: if, say, $c = 0.6$ and $x = 0.8$, then we should conciliate (assuming $x \neq \mu$). Now that cannot happen.[21]

Neither complication is serious. To dispose of the first, we simply choose *two* values of λ before our epistemic interaction: one is to be used in the regular case (in which the confidences we receive are temperate) and the other is to be used if we face an excessively confident interlocutor. Formalism aside, this is probably good practice given the qualitative concerns about excessively confident peers raised here and the real-world prevalence of this behavior.

More generally, λ need not be modeled as two fixed parameters (one for the case of temperate confidence and one for the case of excessive confidence). Rather, λ can be a function of our interlocutor's reported confidence: $\lambda = \lambda(x)$. In this case, equation (3) becomes

$$f'_3(c,x) = c + \lambda(x)(x - \mu),$$

[20]The parameter λ must satisfy "consistency conditions" to ensure a sensical coherence between c, μ, and λ; namely, $\max\left\{\frac{c}{\mu-1}, \frac{c-1}{\mu}\right\} \leq \lambda \leq \min\left\{\frac{c}{\mu}, \frac{1-c}{1-\mu}\right\}$.

[21]The problem is general and does not require strict Steadfastness in the face of excessive confidence. For example, take $c = 0.6$ and $\mu = 0.8$, and suppose we judge $f_3(0.6,1) = 0.65$. That is, we think we should conciliate a little when our interlocutor expresses perfect certainty. Now we have fixed $\lambda = 0.25$, which enables us to compute, for example, $f_3(0.6,0.9) = 0.63$. That is, we conciliate more with the extreme interlocutor than the temperate (i.e., more rational) one.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

which is now nonlinear in x in general. Independence concerns are still avoided, though, provided that $\lambda(x)$ is specified *ex ante*. Not every such function will be admissible; $\lambda(x)$ must be chosen so that the resulting posteriors are probabilistically coherent.

Second, West and Crosse's suggestion for choosing λ can still be used for the majority of cases, involving temperate confidences. And we can still use very high and very low (though not excessive) values of x. Then we can analyze how that value ought to change in the case of excessive confidence and take an appropriate haircut. That is, for cases of excessive confidence, we can choose a (lower) λ, sensitive not just to the fact that x is excessive, but to the degree of excessiveness as well.

Suppose we decide that, in some context, confidences < 0.1 and > 0.9 are excessive. We have a confidence of 0.6 in the proposition at issue and expect our interlocutor to agree with us ($\mu = 0.6$). We reckon that if our interlocutor says that his confidence is in fact 0.9, then the rational post-disagreement confidence is 0.75. We have now fixed $\lambda = 0.5$, and this value can be used for any non-excessive confidence, $0.1 \leq x \leq 0.9$, that we hear.

And we can use lower values of λ for excessive confidence. Perhaps we think that λ should be reduced by 0.02 for confidences of 0.09 and 0.91, reduced by 0.04 for confidences of 0.08 and 0.92, and so on, with λ bottoming out at 0.3 (for confidences of 0 and 1).[22]

We have now selected λ before any confidences are revealed to us. And we can take the firm line —*if contextually appropriate*—that the most excessive confidences call for Steadfastness on our part, while not locking us into that result if a temperate confidence is in fact uttered.

Before explaining how this Bayesian approach deals—effectively—with the objections of Section 3, I want to stress how important the issue of interdependence is, and that care must be exercised in accounting for this possibility.[23] When it comes to real-world disagreement, dependence is the rule rather than the exception. There may be dependence between interlocutors, dependence between yourself and one or more of the interlocutors, or both. Many disagreement norms (e.g., the Equal Weight View) do not contend with the important nuance of this issue for the simple (and in my view fatal) reason that they are unable to incorporate dependence in the first place.

In the Bayesian model discussed, dependence is handled *via* the λ parameter. It would take us too far afield to discuss the modeling of dependence in detail. Observe, though, that in this model, one may take two approaches to incorporating interdependence. One may take a "top down" approach, explicitly considering the details of the scenario (asking and answering questions like: "What epistemic capacities are the same?"; "What common training have we received?"; "What evidence do we share?") and then selecting appropriate λs. In some cases, this may be relatively easy; in others, onerous.

Or, one may take a "bottom up" approach, considering what the appropriate post-disagreement confidence should be given such-and-such simple, fixed, and tractable inputs. That contemplation will yield λs that both appropriately incorporate interdependence and reveal its extent.

Let us move on to the two objections of Section 3. Both involve multiple interlocutors and the appropriate modification of equation (3) is (Genest & Schervish 1985: 1207).[24]

$$f_4(c, x_1, \ldots, x_n) = c + \sum_{i=1}^{n} \lambda_i (x_i - \mu_i). \tag{9}$$

[22]It is necessary to verify that all adjustments maintain compliance with the consistency conditions. In this case, they do maintain compliance.

[23]For commentary on the problem of interdependence within the epistemology of disagreement see, for example, Barnett, 2019, Dietrich, 2010, Dietrich & Spiekermann, 2013, Easwaran et al., 2016, Mulligan, 2021, and Rasmussen et al., 2018.

[24]*N.B.* The published formula contains an error: The c is missing on the right-hand side.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

First, we consider Objection 1. Our two interlocutors are indistinguishable, so we have, preliminarily, $\lambda_1 = \lambda_2$. We compute this parameter as suggested above: by considering what the rational confidence would be if our interlocutors uttered such-and-such. But we do not consider extreme values of x_1 and x_2 for the reasons given, but rather temperate ones.

One example (of many) potential temperate confidences to use would be to consider what we ought to believe if one interlocutor disagrees with us with similar intensity ($x_1 = 0.15$) and one is wholly unsure ($x_2 = 0.5$). For many contexts, a natural thing to say is that equal weight would be appropriate: $f_4(0.85,0.15,0.5) = 0.5$. It is straightforward to compute $\lambda_1 = \lambda_2 = 0.33$.[25]

Now our interlocutors' confidences, $x_1 = 0.2$ and $x_2 = 0.91$, are revealed to us. Before considering the issue of extreme confidence, we can see that $f_4(0.85,0.2,0.91) = 0.66$, which is intuitively sensible and in fact almost the same as equal weighting (0.65).

We are not quite done, though, because x_2 is excessively high. So we take a haircut off λ_2, as described—to, say, 0.30. With rounding, the post-disagreement confidence drops very slightly, to 0.65, and this interlocutor continues to exert significant epistemic influence over us–as is intuitive and theoretically desirable.

Next, Objection 2. Following similar reasoning to the above yields, preliminarily, $\lambda_1 = \cdots = \lambda_5 = 0.17$ (i.e., if it ends up that one person agrees with us, assigning a probability of 0.2, two people disagree, and two people are wholly unsure, then we ought to split the difference, $f_4(0.2,0.2,0.8,0.8,0.5,0.5) = 0.5$).

Now all five reveal to us their (slightly excessively high) shared confidence, 0.91. We take our haircut and obtain $\lambda_1 = \cdots = \lambda_5 = 0.15$. This yields a post-disagreement confidence of 0.73, which is again sensible: Our initial confidence in the proposition was 0.2, and we met five people with (slightly) excessive confidences of 0.91. Qualitatively, we go from being highly skeptical of the proposition to being confident in it. (In this case, equal weighting yields 0.79.)

So the Bayesian approach surmounts the two objections raised in Section 3. It also offers an intuitively compelling motivation for Conciliationism, along with flexibility, via the μ and λ parameters, in grappling with the critical complexities of real-world disagreement (the most important being gradations in competence and ubiquitous interdependence). And, relevant here, it deals with excessive confidence sensibly.

5. Lessons for the Disagreement Debate

We have diagnosed a problem: Excessively confident interlocutors (epistemic peers, inferiors, and superiors) are currently taken more seriously, from the point of view of the theory of disagreement, than they ought to be. I have argued that a Bayesian approach to disagreement, operationalized in equations (3) and (9), is the cure. I now want to offer some broader comments about why the formal Bayesian approach that has been elaborated in this article is so attractive within the philosophical setting of the epistemology of disagreement.[26]

Before I do, though, I want to address Benjamin Anders Levinstein's (2015) apparently contrary claim that "for reliable agents, weight should tend to grow with opinionation" (p. 9). For the reasons given in this article, that cannot be categorically correct (it could, though, be true up to extreme confidences).

[25]Keep in mind that the μ_is must be modeled as well. In the absence of information to the contrary, we make the natural choice that $\mu_1 = \mu_2 = c = 0.85$ (*cf.* Mulligan, 2021). When this is the case, and the λs are nonnegative, equation (9) simplifies to linear pooling (Stone, 1961). But it is easy to consider alternative contexts. For instance, suppose your second interlocutor is a reserved, careful type who tends not to have strong opinions. Then maybe we select $\mu_2 = 0.6$ (in the expectation that he agrees with us, just not as intensely). In such a case, if we feel that $\lambda_1 \neq \lambda_2$, it is necessary to choose two posterior probabilities, going through the procedure twice, to fix the λs.

[26]Mulligan (2021) offers further comments along these lines.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

Levinstein's focus is not what concerns us here; namely, how to rationally update our beliefs in light of disagreement (what Levinstein calls the "micro-theory" of disagreement). Levinstein's argument, rather, putatively *constrains* our updating procedures. It would take us too far afield to analyze his argument in detail, but consider the following case, which he calls **Sunrise**:

> I consider Tom–and indeed, nearly everybody I ever come in contact with–roughly a peer when it comes to the question of whether the sun will rise tomorrow. The reason I hold such an egalitarian view is almost entirely based on my expectations about what these people think. My distribution $\mathfrak{b}$(Tom's credence is x) over potential credences he may have in the proposition that the sun will rise tomorrow has nearly all its weight right around 1… . And, conditional on him having a credence right near 1, I assign him very low expected inaccuracy. If he surprises me and tells me that his credence is close to .5, I'll legitimately ignore him (almost) entirely, since reality diverged so extremely from my expectations. (p. 9)

This is incompatible with the Bayesian approach because we Bayesians update our beliefs *only if* others' reports diverge from our expectations. What Levinstein views as a reason to remain steadfast ($x \neq \mu$) is, to my mind, grounds to conciliate.

And while I certainly agree that "nearly everybody I ever come in contact with [is] roughly a peer when it comes to the question of whether the sun will rise tomorrow," this is not "based on my expectations about what these people think." It is because nearly everybody I ever come in contact with is roughly equally "smart" when it comes to the proposition *the Sun will rise tomorrow*, equally experienced in the relevant ways, equally free from relevant bias, and so on. And, critically, there are never contextual features when this question comes up that might provide my interlocutors some independent insight into the phenomenon.

Of course, $x \neq \mu$ is a necessary but not a sufficient condition for conciliating. The reason we give little heed to a typical interlocutor who expresses a relatively low confidence in *the Sun will rise tomorrow* is that, in this context, such interlocutors provide almost no independent insight above and beyond what we already possess. So $\lambda \approx 0$, and we should indeed "ignore him (almost) entirely."

I'm 99.9% sure that the sun will rise tomorrow. I expect my epistemic peer interlocutor will also report a confidence of $\mu = 0.999$. I compute λ by supposing that, if he in fact reports a confidence of $x = 0.5$, my confidence should barely be shaken, perhaps falling only to 0.990. Then we have $\lambda = 0.02$, and, no matter what my interlocutor says, he has little epistemic impact on me. After all, he's almost certainly joking, or he's distracted (and thus not really responding to the question at all), or he's depressed and expressing a general disquiet (again, not really responding to the question), *etc.*[27]

These examples are, I think, not terribly helpful because they are so thin on context—and again, context matters crucially. Consider this case:

> **Sunrise***
> Your neighbor Bob works for the Federal Emergency Management Agency (FEMA). Bob is not particularly bright; he's certainly less knowledgeable than you are on most topics, and he suffers from various common and mild biases. His duties are secretarial, but they do enable him access to high-level discussions.

You had Bob over for dinner last night. He seemed distracted. When you asked him what was wrong, he made some brief and vague comments about his job, about existential risk, and about

[27]This point—about asymmetries in knowledge between one's own epistemic state and one's interlocutor's—comes up all the time in the literature (see, e.g., Lackey's [2010a: 309*f*] discussion of "personal information" and the remarks in Christensen, 2011.) One way of putting it is that, having much more access to your own behavior and psyche than to others' (even if this intrapersonal access is not perfect), you can rule out epistemic problems in yourself more effectively than you can rule them out in others.

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

solar physics. It didn't make a lot of sense. In the middle of dinner, while he was in the bathroom, you saw his cell phone light up. It was a text from "John Doe." You know that's the FEMA Chief of Staff's name. The message said that Bob was urgently needed to attend an interagency meeting. The message noted that the meeting would be classified at the Top Secret level. After returning from the bathroom, Bob saw the message, quickly made an apology, and rushed out.

> This morning, you see Bob while taking out the trash. You ask him how confident he is in *the Sun will rise tomorrow*.

In **Sunrise***, it seems that Bob's your epistemic inferior (being less intelligent, less knowledgeable, and more biased). But here—unlike in **Sunrise**—there are *contextual* features which imply that this interaction is an important and unusual one. Bob brings something epistemically new to the question at hand.

It would be reasonable to reason thus: "Look, it's probably nothing, but it's possible that something's up with the Sun, or our planet, or something, and Bob's become aware of it through his work. Now that I think about it, he'd probably know about that stuff before nearly anybody else. And whatever Bob's defects, he's a stable and straightforward guy. He wouldn't make a joke about something like that. And he's a teetotaler, so his response isn't going to be muddled by drugs or alcohol. I guess it's possible that he could have unknowingly consumed some drug, but that's very unlikely. So let's see. What if he reports a relatively low confidence in *the Sun will rise tomorrow*? Then it's likely he knows something, or thinks he knows something. Of course, even then, it's probably not a serious cause for alarm. But if Bob reports, say, a confidence of 0.75, I should drop my confidence from 0.999 to 0.95."

Now we have fixed $\lambda = 0.2$. When Bob ends up saying that the probability that the Sun will rise tomorrow is only 50%, you reduce your confidence from 0.999 to 0.9. That does not seem unreasonable. Even though "reality diverged so extremely from my expectations" (I was almost sure that my interlocutor's confidence would be right near 1, but in fact it was 0.5), I do *not*, *and should not*, "ignore him (almost) entirely"—*pace* Levinstein.

Anyways, the point is not that that particular reasoning or choice of parameters is correct. The point is that contending, rationally, with each case of disagreement (including even the simplest and least controversial cases) requires thinking carefully about the context in which it occurs and selecting parameters suitable to that context. The main contextual feature we have been considering in this essay is (the possibility of) excessively high or low reported confidence. But one must also assess one's interlocutor's epistemic credentials, attend to the possibility of dependence, and more. It's remarkable that the Bayesian model, containing only two parameters, provides all this flexibility.

Some of Levinstein's findings are correct and compatible with the Bayesian approach. Notably, he points out that "strong opinions require high self-regard. By extension, if you think a bunch of your friends will end up about as accurate as you are, they'd better tend to have pretty strong opinions too, since a moderate credence can only be so accurate. But if you expect them to be opinionated *and* as accurate as you, you also can't expect that they'll generally be too far from your credence." (pp. 13–14).

Recall our consistency conditions for the single-interlocutor case (n. 20). Your opinion c and your judgment about your interlocutor's "accuracy" λ do indeed constrain your expectation about his credence μ.

To illustrate, let us suppose that your "strong opinion" is $c = 0.95$ and that your interlocutor is "about as accurate as you are," $\lambda = 0.5$.

Then we have

$$\max\left\{\frac{0.95}{\mu-1}, \frac{-0.05}{\mu}\right\} \leq 0.5 \leq \min\left\{\frac{0.95}{\mu}, \frac{0.05}{1-\mu}\right\},$$

or $\mu \in [0.9,1)$. So he's indeed not "too far from your credence."

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

Scenarios like **Sunrise** come up frequently in the disagreement literature. Adam Elga, in trying to defend the Equal Weight View against some obvious counterexamples, recommends:

> Think of your state of mind before [disagreement]. We have assumed that, conditional on the two of you disagreeing, you think that your advisor is just as likely as you to be right. But it is also natural to assume that, conditional on the two of you disagreeing *and your finding her answer utterly insane*, you think that you are much more likely to be right. If so, then when that circumstance arises the equal weight view instructs you to favor your own answer… . What makes the above answer work is an asymmetry in the case. You find your advisor's answer insane. (2007: 491)

In these "insane" cases (e.g., **Sunrise**, but not **Sunrise***), one should not significantly modify one's belief (let alone afford one's interlocutor equal weight). And, as we have discussed, that's precisely what occurs, in a nuanced and principled way, in the Bayesian approach (with λ very small).

Note (and this is important for the Bayesian, and arguably different than how this issue has been discussed in the literature to this point)[28] that our interlocutor is "insane" *not* because we are highly confident in our belief, but *because that is contextually appropriate.* Indeed, it is not hard to imagine cases in which we are deeply uncertain about some proposition but yet would disregard any disagreement on grounds of "insanity."[29]

The way Jennifer Lackey (2010b) frames the central problem of the epistemology of disagreement is that the "nonconformist" (i.e., adherent of Steadfastness) cannot grapple with the *One Against Many Problem*, and the "conformist" (adherent of Conciliationism) cannot grapple with the *Many Against One Problem.*

The One Against Many problem is: "If enough epistemic peers disagree with me in an appropriately independent fashion, clearly I should significantly revise my belief" (pp. 278–79). We have seen (Objection 2) how the Bayesian approach solves the problem, and how this caveat of "appropriately independent" may be made precise.

The Many Against One problem is: "If *many* epistemic peers independently agree with one on the answer to a question, yet only *one* epistemic peer disagrees, surely substantial doxastic revision is not required; however, it is not clear how [Conciliationism] will provide a principled explanation of this." I hope it is clear that, in such a case, equation (9) will protect one's initial belief, and the sole dissenting peer will not substantially affect it.

That said, if we are to take Lackey's scenario literally—if these many peers are truly "independent"—then substantial doxastic revision *may well* be required, but *away* from the dissenter, not toward him. For instance, if one has a pre-disagreement confidence of 0.7, and one receives reports from five independent peers, one of whom reports a confidence of 0.4 and the other four of whom report a confidence of 0.7, then one's post-disagreement confidence may be >0.7! (This phenomenon is sometimes known as "risky shift".)

And Christensen has lamented that we are "a long way from offering a detailed conciliationist recipe for accommodating the evidence provided by the disagreement of others" (2011: 17). He says that what is needed is a "general Conciliatory principle" whose prescriptions are sensitive to "the likelihood that [one's interlocutor's] expressed disagreement is sincere," "her degree of well-informedness," and "the likelihood of her having reasoned correctly from the evidence she has" (p. 17). We have seen that the Bayesian approach is such a principle, satisfying these, and indeed other, essential desiderata.

[28]See, for example, Lackey (2010b: 285*f*).

[29]Example: A coin flip's coming up. The proposition in question is *the flip will be heads.* You and your interlocutor have been given very strong evidence (like the results of a recent laboratory analysis) that the coin is fair. Your confidence in the proposition is 0.5, and—assuming no other relevant contextual information—anything other than that is, in Elga's terminology, "insane."

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

Even with such a principle, Christensen says that

> many questions remain open: How far, in general, should one revise one's belief? How does the principle extend to cases where initial confidence is distributed differently–*e.g.*, where both parties are more confident that P than ∼P but to different degrees? … The initially attractive idea of uniformly splitting the difference in credences does not sit well with the motivations for Conciliationism. So we are a long way from having a formula, or even a recipe with quantities, describing in general how one should react to disagreement. (p. 17)

None of this is aspirational. We have seen that there is a simple and intuitive way to handle confidences that are "distributed differently" and that "uniformly splitting the difference" (i.e., the Equal Weight View) is appropriate only in special, usually highly stylized, cases. Equation (9) is the formula, the "recipe with quantities" desired.

6. Conclusion

A salient, and to me distressing, part of contemporary life is the certainty that people have in their beliefs. This is true even (especially?) when these beliefs concern difficult matters of morality and politics. Whatever other deleterious effects excessive confidences have, they are dangerous from the epistemic point of view. We must take care not to let them affect our beliefs too much. This may be accomplished by modifying our conciliatory norms as suggested. It is a small change, but an important one.

Thomas Mulligan is a Visiting Scholar at Georgetown University and an Adjunct Researcher at the RAND Corporation. His research has spanned political philosophy, epistemology, decision theory, and the philosophy of artificial intelligence. He is the author of Justice and the Meritocratic State (2018, Routledge) and is currently working on the political economy of AI.

References

Barnett, Z. (2019). Belief dependence: How do the numbers count? *Philosophical Studies*, 176(2), 297–319. https://doi.org/10.1007/s11098-017-1016-0

Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review*, 116(2), 187–217. https://doi.org/10.1215/00318108-2006-035

Christensen, D. (2011). Disagreement, question-begging and epistemic self-criticism. *Philosophers' Imprint*, 11, 1–22.

Dietrich, F. (2010). Bayesian group belief. *Social Choice and Welfare*, 35(4), 595–626. https://doi.org/10.1007/s00355-010-0453-x

Dietrich, F., & Spiekermann, K. (2013). Independent opinions? On the causal foundations of belief formation and jury theorems. *Mind*, 122(487), 655–685. https://doi.org/10.1093/mind/fzt074

Easwaran, K., Fenton-Glynn, L., Hitchcock, C., & Velasco, J. D. (2016). Updating on the credences of others: Disagreement, agreement, and synergy. *Philosophers' Imprint*, 16, 1–39.

Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3), 478–502. https://doi.org/10.1111/j.1468-0068.2007.00656.x

Feldman, R. (2007). Reasonable religious disagreements. In L. M. Antony (Ed.), *Philosophers without gods: Meditations on atheism and the secular life* (pp. 194–214). Oxford University Press. https://doi.org/10.1093/oso/9780195173079.003.0016

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Francese, C., & Richeson, D. (2007). The flaw in Euler's proof of his polyhedral formula. *American Mathematical Monthly*, 114 (4), 286–296. https://doi.org/10.1080/00029890.2007.11920417

Gardiner, G. (2014). The commutativity of evidence: A problem for conciliatory views of peer disagreement. *Episteme*, 11(1), 83–95. https://doi.org/10.1017/epi.2013.42

Gelfert, A. (2011). Who is an epistemic peer? *Logos & Episteme*, 2(4), 507–514. https://doi.org/10.5840/logos-episteme2011242

Genest, C., & Schervish, M. J. (1985). Modeling expert judgments for Bayesian updating. *Annals of Statistics*, 13(3), 1198–1212. https://doi.org/10.1214/aos/1176349664

Jehle, D., & Fitelson, B. (2009). What is the "equal weight view"? *Episteme*, 6(3), 280–293. https://doi.org/10.3366/E1742360009000719

Kelly, T. (2005). The epistemic significance of disagreement. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 1, pp. 167–196). Oxford University Press.

Kuan, K.-H. (2021). Beyond linear conciliation. *Synthese*, 198(12), 11483–11504. https://doi.org/10.1007/s11229-020-02809-7

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press

Lackey, J. (2010a). A justificationist view of disagreement's epistemic significance. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Social epistemology* (pp. 298–325). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199577477.003.0015

Lackey, J. (2010b). What should we do when we disagree? In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 3, pp. 274–293). Oxford University Press.

Levinstein, B. A. (2015). With all due respect: The macro-epistemology of disagreement. *Philosophers' Imprint*, 15, 1–20.

Lindley, D. V. (1971). *Making decisions*. John Wiley & Sons Ltd.

Lougheed, K. (2020). *The epistemic benefits of disagreement*. Springer.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Mulligan, T. (2021). The epistemology of disagreement: Why not Bayesianism? *Episteme*, 18(4), 587–602. https://doi.org/10.1017/epi.2019.28

Plous, S. (1993). *The psychology of judgment and decision making*. Temple University Press.

Rasmussen, M. S., Steglich-Petersen, A., & Bjerring, J. C. (2018). A higher-order approach to disagreement. *Episteme*, 15(1), 80–100. https://doi.org/10.1017/epi.2016.43

Stone, M. (1961). The opinion pool. *Annals of Mathematical Statistics*, 32(4), 1339–1342. https://doi.org/10.1214/aoms/1177704873

Taylor, R., & Wiles, A. (1995). Ring-theoretic properties of certain Hecke algebras. *Annals of Mathematics*, 141(3), 553–572. https://doi.org/10.2307/2118560

West, M., & Crosse, J. (1992). Modelling probabilistic agent opinion. *Journal of the Royal Statistical Society B*, 54(1), 285–299. https://doi.org/10.1111/j.2517-6161.1992.tb01882.x

Wiles, A. (1995). Modular elliptic curves and Fermat's last theorem. *Annals of Mathematics*, 141(3), 443–551. https://doi.org/10.2307/2118559

Cite this article: Mulligan, T. 2025. The Rational Response to Excessive Confidence. *Canadian Journal of Philosophy* 55: 452–468, doi:10.1017/can.2026.10070

https://doi.org/10.1017/can.2026.10070 Published online by Cambridge University Press