The Epistemology of Disagreement: Why Not Bayesianism?

(penultimate draft—forthcoming in *Episteme*)

Thomas Mulligan

Georgetown University

Put people together and before long you will find them at odds. It appears to be a fact of our species that we have disagreed, do disagree, and will disagree—and about almost anything. We disagree about important issues of morality and politics; we disagree about sports and other banalities. Our disagreements are sometimes silly, but they are sometimes sober and reasonable, at least *prima facie*: None of us is obviously irrational, no argument beyond the pale.

Philosophers have dutifully attended to the phenomenon of disagreement, considering, in particular, its epistemology.[1] The central question in the epistemology of disagreement debate is how a rational person modifies her belief in a proposition *X* (if at all) when she learns that someone else disagrees about *X*. The ultimate reason for such revision, if indeed it is warranted, may be purely intellectual—one wishes to have a maximally justified belief for its own sake. More frequently, though, it is to motivate action: "The goal of maximally justified belief . . . is primarily a goal of individuals who need to act"

[1] Philosophers have also worked on (1) disagreement's ramifications for political authority (*e.g.* Rawls 1993); (2) *social choice theory*—the aggregation of individual preferences into a group preference (which Arrow (1950) showed to be intractable); and (3) *collective decision-making*—how individuals who share a common goal but disagree about how to pursue that goal ought to comport themselves (the most famous result in collective decision-making is the Marquis de Condorcet's *jury theorem* (1785)).

(Everett 2015: 278). One wishes to form a maximally justified belief about what 20% of the restaurant bill is (Christensen 2007) to leave a proper tip.

Although an area of active research for only about a decade,[2] the epistemology of disagreement has attracted intense interest—a result, perhaps, of our hyper-partisan political climate, and worry about its social effects.[3]

The method used by philosophers investigating disagreement today, and the method which underpins the foundational work (n. 4), is to argue for a disagreement norm *a priori*. For example, this is how David Christensen makes his case for conciliationism:[4]

> Disagreement gives one evidence that one has made a mistake in interpreting the original evidence . . . Thus the persistence of the degree of disagreement on important issues . . . indicates that, in general, practitioners in the field do not form beliefs reliably. If one is a practitioner in such a field, then, absent some reason to think oneself special, one should not have confident opinions on the field's controversial questions. (2007: 757)

On the other side of the debate, we have Thomas Kelly's reasoning in favor of steadfastness:

> The question of how well someone has evaluated the evidence with respect to a given question is certainly the kind of consideration that is relevant to deciding whether his or her judgement ought to be credited with respect to that question. That is, it is exactly the sort of consideration that is capable of producing the kind of asymmetry that would justify privileging one of the two parties to the dispute over the other party. And from my vantage point—as one of the parties *within* the dispute, as opposed to some on-looking third party—it is just this undeniably relevant difference that divides us on this particular occasion. (2005: 179)

---

[2] One can find harbingers of the contemporary disagreement debate in Lehrer and Wagner 1981 and Loewer and Laddaga 1985.

[3] Seminal work in the epistemology of disagreement includes Christensen 2007, Elga 2007, Kelly 2005, Lackey 2010a, and van Inwagen 1996. The best introduction to disagreement is Feldman 2007 (see also Christensen 2009).

[4] Conciliationists hold that one ought to revise one's belief in the face of disagreement from a suitable epistemic interlocutor. See, *e.g.*, Christensen 2007, Elga 2007, and Feldman 2007. Adherents to steadfastness believe, to the contrary, that one may "stick to one's guns" epistemically, maintaining one's original confidence in the proposition at dispute. For defenses of steadfastness, see, *e.g.*, Bergmann 2009, Kelly 2005, and van Inwagen 2010.

Christensen and Kelly go on to adduce examples from real life which are supposed to show that their preferred norms are correct.

I am convinced that this approach is misguided. We should not endorse a disagreement norm on *a priori* grounds and a handful of intuitive cases, and then impose it on those cases for which intuitions go the opposite way. Rather, we should aspire to a principled approach to belief revision that yields steadfast norms in appropriate cases and conciliatory norms in appropriate cases.

The purpose of this paper is twofold. First, I argue that such a principled approach is possible—if we apply neglected tools from Bayesian analysis.

Second, I show that the Bayesian modelling strategy I commend satisfies three vital desiderata which mainstream approaches to disagreement, conciliatory and steadfast, do not. To wit, Bayesian modelling can (1) deal with multiple epistemic interlocutors; (2) allow one to suitably modify one's beliefs in the face of disagreement from epistemic *superiors* and *inferiors* (*i.e.* not just epistemic *peers*); and (3) accommodate the possibility of dependence between interlocutors.

I have organized this paper as follows. In §1, I define the term *epistemic peer* and provide necessary notation. §2 introduces the Bayesian approach to belief revision. §3 argues that dependence among epistemic interlocutors is not just ubiquitous in the real-world but critically important from a formal point-of-view. Our norms must be capable of accommodating it. I also critique a recent effort of Easwaran *et al.* (2016) to provide a disagreement norm in the Bayesian spirit. §4 provides what I believe to be a better model for disagreement. I conclude in §5.

1. The epistemology of disagreement: concepts and modelling assumptions

I shall not give an overview of the epistemology of disagreement debate (for that, see the references listed in n. 4). In this section, I only want to define a term and provide some notation.

The term is "epistemic peer", first coined by Gary Gutting (1982). Intuitively, if I believe that Elizabeth Woodville was the wife of Henry VI, and my 8-year-old cousin disagrees with me about this, I need not lose confidence in my belief—for I am much more likely than he is to be correct about this historical fact. On the other hand, when a professor of British history tells me that I am wrong, I certainly must lose confidence. But what is the rational response when someone as likely as I am to be correct about *Elizabeth Woodville was the wife of Henry VI* disagrees with me?[5] Most epistemologists believe that this the interesting case—disagreement with an *epistemic peer*—and it has been the focus of the debate.

Kelly defines the term thus:

> 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)

This is a typical definition (*cf.* Gelfert 2011 and Matheson 2015), and it will suffice for our purposes.

Of course, that disagreement with peers is of epistemic interest does not imply that disagreement with non-peers is not. Yet little attention has been paid to how one should

[5] Woodville was in fact the wife of Edward IV—not Henry VI.

revise one's beliefs in light of disagreement from an epistemic superior or an epistemic inferior. And the attention that has been paid to those cases (*e.g.* Zagzebski 2012) tends to focus on special contexts, like morality. The implicit assumption is that belief revision is warranted when one comes into contact with an epistemic superior, and unwarranted when one comes into contact with an epistemic inferior.

This will not do. Set aside that is a rare thing to interact with people whom we can honestly say are pure epistemic peers. All our epistemic interactions take place within a complicated nexus of inferiors, peers, and superiors. The "smartest" person we know may believe X; two marginally less smart people may believe $\sim X$; two peers may believe X while one believes $\sim X$; and all the while four inferiors believe $\sim X$. What to do, epistemically? As things stand now, no guidance is forthcoming.

What we, epistemologists interested in disagreement, should like to have is an approach to disagreement that incorporates not only gradations in confidence, which our norms already do, but also gradations in *competence*, which our norms do not.

I define some terms. The goal of the epistemology of disagreement debate is to identify the correct *disagreement norm*. We shall be considering a finite set of people, $\boldsymbol{V} = \{v_1, v_2, \ldots, v_n\}$, who have opinions about some proposition X. Although philosophers typically consider only the special case of $n = 2$,[6] we shall not so limit ourselves.

Associated with each v_i is a *confidence*, $c_i \in (0, 1)$. We interpret a person's confidence in X as follows: As c_i approaches 1 (0), v_i approaches certainty that X is true (false). Although within the philosophical literature one more commonly sees $c_i \in [0, 1]$, it is better to use the open interval. This becomes relevant for technical reasons later on, but

[6] Exceptions include Gardiner (2014) and Mulligan (2015).

it also makes more sense from the Bayesian point-of-view. To say that one has a confidence of 1 (0) is to say that no future evidence could shake one's belief that the proposition at issue is true (false). That is wrong even in the strongest real-world contexts. Even our beliefs about putative necessary truths might one day be undermined. Sometimes we discover that a mathematical "proof" we thought was rigorous is in fact subtly flawed.

Let us suppose, without loss of generality, that v_1 is the one trying to decide whether or not to modify her belief in X in light of disagreement.

2. Bayesian belief revision

It is surprising to me that little effort has been devoted to tackling the core problem of the epistemology of disagreement—how to revise one's belief given the beliefs of others—with tools from Bayesian analysis. I am aware of only two broad attempts along these lines in the philosophical literature. Typically, the relevant papers argue, often convincingly, that some disagreement norm is incompatible with a Bayesian principle or its overarching philosophy. The goal of this paper, in contrast, is to provide a *general* Bayesian solution to the disagreement problem.

First, there are a number of persuasive arguments that concilationism, usually understood in its "equal weight" form (the idea, roughly, that when two epistemic peers disagree, the rational thing for them to do is to "meet in the middle") is incompatible with Bayesianism—because, for example, it violates conditionalization. (*Cf.* Isaacs 2019, Jehle and Fitelson 2009, Lasonen-Aarnio 2013, Levinstein 2015, and Shogenji Manuscript.)

Second, there is Easwaran *et al.*'s (2016) derivation of a disagreement norm they call "Upco" ("Updating on the credences of others"). This is the best-developed Bayesian approach to disagreement in the epistemological literature, and I shall consider it in some detail in the next section.

As we shall see, not only does a proper Bayesian strategy provide a means for updating confidences given others' confidences (indeed, this is Bayesianism's *raison d'être*), it satisfies the three desiderata mentioned in the introduction: It deals with multiple epistemic interlocutors; provides guidance for updating beliefs given disagreement from interlocutors of whatever competences; and it ensures that facts about epistemic dependence influence beliefs appropriately.

Moreover, the Bayesian strategy provides steadfast norms in those cases in which, intuitively, it is rational to "stick to your guns". And it provides conciliatory norms for those cases in which belief revision seems right. For a good Bayesian, sometimes steadfasters like Kelly are correct; other times, conciliationists like Christensen are on the right side of things. We should not impose either norm on scenarios for which it is inappropriate—even though philosophers often try to do just that.

Now, some epistemologists, like Richard Feldman (2009), have argued informally against a "one size fits all" approach to disagreement, often as part of a "total evidence" approach:

> I am not endorsing universal principles asserting that it is never reasonable to maintain one's belief, I am arguing that evidence of peer disagreement is evidence against one's original belief. It is consistent with this that, in many cases, it is strong evidence against one's original belief, strong enough to render that belief no longer justified. (Feldman 2009: 304)

And Kelly (2010) discusses coming to terms with multiple epistemic interlocutors and dependence through a total evidence approach. For adherents of this approach, perhaps this paper, and the Bayesian paradigm more broadly, can provide a useful formal framework for determining how one's total evidence should bear on a given hypothesis.

Our key move will be for v_1 to regard her interlocutors' judgments as random variables, the values of which—namely, $c_2, \ldots, c_n$—are revealed to her by $v_2, \ldots, v_n$. Then, v_1 treats $c_2, \ldots, c_n$ as data relevant to X on which she can update c_1. Obviously this is a very different methodology than the typical, *a priori* approach described in the introduction. But it is also different than Easwaran *et al.*'s Upco, which involves no probabilistic modelling at all, but is, rather, a function which falls out as a special case of Bayes's Law (under the assumption that epistemic interlocutors are independent conditional on X).[7]

We begin by noting that confidence is typically interpreted as subjective probability (*cf.* §1):

$$c_i = \Pr_i(X) \tag{2.1}$$

Note that since X and $\sim X$ are mutually exclusive and jointly exhaustive of the sample space (the proposition is either true or it is false, and not both), $(1 - c_i)$ is v_i's subjective probability that $\sim X$.

Next, consider the special case of $n = 2$.[8] v_1 and her interlocutor, v_2, are considering X (*e.g. the defendant is guilty*). v_2 reports a confidence of c_2 in X. Denoting v_1's post-

[7] See Morris 1974 for a derivation and discussion of Upco.

[8] The foundational work here, underlying what follows, was done by Morris (1974).

disagreement confidence by c'_1, Bayes's Law provides unambiguous guidance to v_1 about how, rationally, she should proceed:[9]

$$c'_1 = \Pr(X \mid c_2) = c_1 \times \frac{\Pr(c_2 \mid X)}{c_1 \times \Pr(c_2 \mid X) + (1 - c_1) \times \Pr(c_2 \mid {\sim}X)} \tag{2.2}$$

Notice how the disagreement problem reduces to specification of the likelihoods $\Pr(c_2 \mid X)$ and $\Pr(c_2 \mid {\sim}X)$. That is, to reach a maximally justified belief, v_1 must answer two questions: (1) "What is the probability that my interlocutor would say what he did (*viz.* c_2) if the state of the world is such that X (*e.g.* the defendant is in fact guilty)?" And (2) "What is the probability that my interlocutor would say what he did if the state of the world is such that ${\sim}X$ (the defendant is innocent)?"

One can see immediately how the Bayesian approach satisfies the desideratum of gradations in competence (which, again, dominant epistemological approaches do not); these are incorporated into the likelihood functions themselves.

For example, take the special case in which v_1's interlocutor is not only her epistemic superior but is epistemically infallible (and v_1 knows this): v_2 reports $c_2 = 0$ if X is false, $c_2 = 1$ if X is true, and nothing else. Then if v_1 hears "0" from v_2, $c'_1 = 0$. (Because $\Pr(c_2 = 0 \mid X) = 0$.) If v_1 hears "1" from v_2, then $c'_1 = 1$. (Because the second term on the RHS of equation (2.2) becomes $\frac{1}{c_1}$.) Of course, in general v_1 will have to specify likelihood functions that cover the entire domain of c_2—from 0 to 1. But the principle is the same.

[9] I suppress the subscript on the probability function here on out. I am also going to abuse notation a little, using c_i to refer both to the realization of a random variable and to the random variable itself.

Another alluring feature of this approach is that it satisfies the second desideratum: It easily generalizes to *n* of arbitrary size. Again, by Bayes's Law:

$$c'_1 = \Pr(X \mid c_2, \ldots, c_n) = c_1 \times \frac{\Pr(c_2, \ldots, c_n \mid X)}{c_1 \times \Pr(c_2, \ldots, c_n \mid X) + (1 - c_1) \times \Pr(c_2, \ldots, c_n \mid {\sim}X)} \quad (2.3)$$

The likelihoods can be put into more manageable form. By the definition of conditional probability, Pr(A, B | C) = Pr(A | B, C) × Pr(B | C). Thus,

$$\Pr(c_2, \ldots, c_n \mid X) = \Pr(c_n \mid c_2, \ldots, c_{n-1}, X) \times \Pr(c_2, \ldots, c_{n-1} \mid X) \quad (2.4)$$

and

$$\Pr(c_2, \ldots, c_n \mid {\sim}X) = \Pr(c_n \mid c_2, \ldots, c_{n-1}, {\sim}X) \times \Pr(c_2, \ldots, c_{n-1} \mid {\sim}X) \quad (2.5)$$

The second terms on the right-hand-sides of equations (2.4) and (2.5) can be expanded in a similar way. Doing that, and substituting into equation (2.3), yields:

$$c'_1 = c_1 \times \frac{\prod_{i=2}^{n} \Pr(c_i \mid c_2, \ldots, c_{i-1}, X)}{c_1 \times \prod_{i=2}^{n} \Pr(c_i \mid c_2, \ldots, c_{i-1}, X) + (1 - c_1) \times \prod_{i=2}^{n} \Pr(c_i \mid c_2, \ldots, c_{i-1}, {\sim}X)} \quad (2.6)$$

Again the disagreement problem is one of specifying likelihoods. Now, especially when it comes to multiple interlocutors, this may be an onerous task. It requires that v_1 detail her interlocutors' intelligence, susceptibility to bias, and other epistemic features. As a result, in the 1970s and 1980s, decision theorists proposed a number of coarse-grained models for real-world use.[10] It seems that philosophers are unaware of this body of work, despite the relevance for the epistemology of disagreement suggested by some of its titles (*e.g.* French's "Updating of Belief in the Light of Someone Else's Opinion"). I want to describe one such model here, to illustrate the applicability of the Bayesian modelling approach to our contemporary disagreement debate.[11]

The first model was developed by Peter Morris (1983) and Robert Winkler (1968). I have chosen it because it is simple and because it incorporates the two desiderata just described.

The idea underlying the model is to treat individuals' beliefs about some event, like a defendant's guilt, as Beta-distributed random variables. The reported c_is are regarded as the means of those distributions. The beta distribution is appropriate because we are seeking to represent a distribution of probabilities. And it yields a new, post-disagreement distribution by summing over parameters that define the individual distributions. The mean of that new distribution may then be adopted as the post-disagreement confidence.

This provides the following disagreement norm (I omit the derivation here, it may be found in the cited work):

$$c'_1 = \sum_{i=1}^{n} w_i c_i \tag{2.7}$$

[10] Good summaries of this literature may be found in Clemen and Winkler 1990 and 1999 and French 1985.

[11] Other models, not described here, include those of Clemen and Winkler (1987), French (1981), and Lindley (1985).

where

$$\sum_{i=1}^{n} w_i = 1 \tag{2.8}$$

This is a simple weighted average, where the opinions of the v_is are granted influence in accordance with v_1's view of their relative competence. In the special case in which v_1 regards them all as epistemic peers, weights are set to $\frac{1}{n}$.

Note two things. First, our two desiderata are incorporated—the w_is provide for differences in competence, and as many interlocutors as v_1 likes may offer their opinions on X for v_1's consideration. Second, this norm is essentially the same as conciliationism's equal weight view, albeit more general.[12]

An example of the norm in action may be helpful. Consider the "complicated nexus" problem described in §1: v_1 is trying to form a maximally justified belief about X in light of disagreement from 10 epistemic interlocutors—some peers, some superiors, and some inferiors.

Under Morris and Winkler's model, v_1 ought to do two things: (1) obtain reports from $v_2, \ldots, v_{11}$ regarding their confidences in X; and (2) assess the relative competences of $v_1, \ldots, v_{11}$. Suppose that this yields:

---

[12] It is also essentially the same as the (non-Bayesian) "linear opinion pool" (see, *e.g.*, Stone 1961).

Epistemic agent	Reported confidence	Relative competence
v_1	0.5	0.1
v_2	0.7	0.16
v_3, v_4	0.3	0.14
v_5, v_6	0.8	0.1
v_7	0.4	0.1
v_8, v_9, v_{10}, v_{11}	0.2	0.04

Then $c'_1 = 0.48$. The computation is straightforward, as is the solicitation of confidence information from $v_2, \ldots, v_{11}$. The only challenge for v_1 is assessing relative competence.

Now this norm has a serious drawback, a drawback which plagues theories in the epistemology of disagreement but which has hardly been attended to in the disagreement literature.[13] Namely, it implicitly endorses the idea that if there is no disagreement to begin with (*i.e.* if all epistemic interlocutors share the same confidence), then the post-disagreement confidence should simply equal the shared, pre-disagreement confidence. This is sometimes known as the "unanimity condition", but it is a bug, not a feature, of a theory, even in the special case of agreement with epistemic peers.

Whether the unanimity condition should hold or not turns on whether there exists dependence between (1) our epistemic agent and her interlocutors, and (2) the interlocutors themselves. For reasons I shall now give, any viable disagreement norm must be capable of modelling such dependence.

[13] But see Barnett Forthcoming, Dietrich 2010, and Easwaran *et al.* 2016.

3. Disagreement and dependence

v_1, v_2, and v_3 are professional horseplayers, trying to form a maximally justified belief in *Judy's Lightning will win the race*. They regard each other as epistemic peers, and they have good evidence that they are in fact peers: They've been betting on races for a long time, and have had the same success in picking winners.

But v_1, v_2, and v_3 are not identical. In particular, v_1 and v_2 share the same handicapping methodology: They rely on how horses appear the morning of the race. v_3, on the other hand, has developed a mathematical system for predicting winners on the basis of diverse historical data. Nevertheless, these two methodologies appear equally good; v_1, v_2, and v_3 win with equal frequency. Naturally, v_1 and v_2 tend to win together, because they share the same methodology. In contrast, v_3 sometimes wins when v_1 and v_2 lose (and *vice versa*).

Suppose that, before this race, each reports the same confidence, γ, in *Judy's Lightning will win the race*. According to standard disagreement theory, v_1 should not change her confidence in this proposition (obviously γ is unmodified under steadfastness, and it implicitly stays the same under most variants of conciliationism, too—the arithmetic average of $\{\gamma, \gamma, \gamma\}$ is γ).[14]

Two questions to consider: (1) Should v_1's confidence in *Judy's Lightning will win the race* be unchanged by her interaction with v_2 and v_3, given that the three do not disagree about the probability of this event? (2) If v_1's post-"disagreement" confidence

[14] Although see Elga 2013.

should not remain the same, should v_2 and v_3 exert the same epistemic influence on v_1 when she modifies her judgment?

The answer to (1) is, *pace* current theory, "no". Even though v_3 is an epistemic peer, equally good at getting to the truth of *Judy's Lightning will win the race*, v_3 is *different* than v_1. And that difference means that their assessments are at least partially independent. If she is rational, v_1 will recognize that independence and use it appropriately to modify her confidence. For example, if $\gamma = 0.8$, then v_1's post-"disagreement" confidence will be greater than 0.8. The knowledge that a different handicapping methodology, even if no better than your own, is also highly confident that it has picked a winner provides you with greater reason to believe that you have got things right.

As far as (2) is concerned, v_2 and v_3 should certainly not exert the same epistemic influence over v_1. Indeed, because v_2 is more-or-less a copy of v_1, c_2 is not a useful datum when $c_1 = c_2 = 0.8$. That is precisely what v_1 expects to hear prior to her interaction with v_2, and so conditionalizing upon it should not affect her prior judgment. Because there is perfect dependence between v_1's judgment and v_2's judgment, and v_1 knows this, there is nothing to be gained epistemically through interaction with v_2 in this case.

The possibility of dependence between epistemic agents illuminates the limitations of Easwaran *et al.*'s Upco:

$$c'_1 = \frac{c_1 \times c_2}{c_1 \times c_2 + (1 - c_1) \times (1 - c_2)} \tag{3.1}$$

While this norm does allow for violations of unanimity (what Easwaran *et al.* call "synergy")—as it sometimes should—it fails to account for dependence, as in the example just given. It would allow v_2 and v_3 to exert the same epistemic influence over v_1. And that, as we have seen, would be a mistake.

Under Upco, if v_1 and v_2 interact when both have a confidence of $\gamma = 0.8$, then the post-"disagreement" confidence is 0.94. The same result is reached if v_1 and v_3 interact. But such a high post-"disagreement" confidence is only plausible in the latter case. Because v_2 provides no independent insight, v_1 should maintain a confidence of 0.8 after interacting with v_2 alone. Upco fails to account for this important difference.

To their credit, Easwaran *et al.* recognize that Upco is limited in this way. But they do not grapple with "the general question of how to deal with peer update when we think there are correlations between one's peers" (p. 31), suggesting, instead, that Upco's use be restricted to scenarios of disagreement involving perfect independence between epistemic agents. But I stress that this is not a minor limitation; it is a loss of generality which renders the norm useless for real-world use. It is a struggle to imagine a real-world scenario in which perfect independence holds. Two philosophers disagree about the morality of some new law? Consider all the common training they receive. Two jurors disagree about a defendant's guilt? Think of the common evidence, presented at trial, on which their judgments rely. Two weathermen disagree about whether it will rain tomorrow? Note that both base their judgments on the same radar data.

Easwaran *et al.* do offer a conjecture about how Upco might handle dependence—but I do not think that it will work. They suggest that we assign to each term in Upco an exponent representing the weight of that interlocutor's opinion, where the weight assigned to the c_1 terms is set to 1, and the weight of "fully independent peers" is likewise 1. Then, if peers' opinions are correlated, we reduce the weights assigned to those peers. For example, if two peers are perfectly correlated, then they should be treated by Upco as *one* "fully independent peer" by assigning each a weight of one-half.

At the same time, the exponents are supposed to handle gradations in competence; indeed, this is why they are introduced in the first place. (Again, Easwaran *et al.* explicitly avoid in-depth analysis of dependence in their paper.) For example, "if we raise [c_i] to the power of 2, we are treating [v_i's] report . . . as equivalent to the report of two independent peers with weight 1 reporting that credence." (Easwaran *et al.* 2016: 30).

Here's the problem. Suppose that I think some proposition is false. My epistemic superior thinks that it is true with confidence c_2. Whatever else we might want to say about my post-disagreement confidence, it should not be greater than c_2. But we can choose values for this Upco variant that delivers such a result:

$$c'_1 = \frac{0.4^1 \times 0.7^2}{0.4^1 \times 0.7^2 + (1-0.4)^1 \times (1-0.7)^2} = 0.78 \qquad (3.2)$$

To deal with dependence, I suggest we look elsewhere.

4. A better way

We ought to model disagreement in a way that explicitly takes into account correlation between epistemic interlocutors. Our models should not hew to unanimity; they should yield a post-disagreement confidence which turns in part on pre-disagreement dependence.

An excellent starting point would be the work of Christian Genest and Mark Schervish (1985), who recognize that even though ideally rational agents should employ equation (2.6), that formula is unhelpful for real-world use. This is because it requires that v_1 specify $\Pr(c_2, \ldots, c_n \mid X)$ and $\Pr(c_2, \ldots, c_n \mid {\sim}X)$, each of which are joint distributions over $n - 1$ random variables.

Genest and Schervish show that so long as (1) v_1 can specify the means of the marginal distributions and (2) plausible consistency conditions hold (more on this below), then a maximally justified belief is given by:

$$c'_1 = c_1 + \sum_{i=2}^{n} \lambda_i (c_i - \mu_i) \tag{4.1}$$

where μ_i is the mean of i's confidence distribution (as specified by v_1), and the λ_is are the coefficients of linear regression of X on the vector $(c_2, \ldots, c_n)$.

To reiterate: c_1 is v_1's pre-disagreement confidence in the proposition at issue. c_i is the stated confidence of v_i in the proposition. v_1 says to herself, for each of her epistemic interlocutors, "My interlocutor might report many possible confidences—from very close to 0 to very close to 1. Some values are more likely than others. What is the mean confidence that I expect to hear?" That is μ_i.

Note that if v_1's interlocutor reports what she expects him to (that is, if $c_i = \mu_i$), then her confidence is unchanged by their interaction. This makes sense from the Bayesian point-of-view; the interlocutor's judgment was already baked into c_1.

Suppose, for example, that the government announces a tax hike on the rich. I have a confidence of $c_1 = 0.8$ that this policy is just. My colleague down the hall is a conservative, so I expect he'll find the policy to be unjust. Perhaps I think he's most likely to report a confidence of 0.2 (if c_2, *qua* random variable, is Normal, the mean equals the mode). Now if I ask him what he thinks about the policy, and he tells me, as I expect he will, that it is unjust, my confidence should be little affected. I already knew that. But if this conservative agrees with me that the tax hike is just—well, that is surprising. It is genuinely new and useful information, and so it gives me grounds to increase my confidence that the policy is a just one.

Of course, technically c_2 cannot be Normal—because the support of the normal distribution is $(-\infty, \infty)$ and c_2 requires a support of (0, 1). One can truncate the normal distribution to (0, 1), but then it is no longer generally true that its mean will equal its mode. Nevertheless, for the purpose of on-the-fly belief revision, v_1 would not go far wrong by imagining c_2 distributed "normalish" on (0, 1), and taking its peak to be μ_2. And, if v_1 desires a precise answer, she can calculate the actual mean for her chosen Truncated Normal on (0, 1).

Note that it will frequently be sensible for v_1 to set $\mu_i = c_1$. That is, v_1 can presume that the mean realization of c_i (*qua* random variable) is what she, v_1, already believes. This is not the case in the above example, because there I know that my opinion about the policy and my colleague's opinion are likely to be opposed given our political differences. But often $\mu_i = c_1$ is sensible. (Example: I'm walking down the street, the sky is looking gloomy, and I think there's a 60% chance of rain. I ask a passerby what he thinks. It's highly unlikely that he'll say "60%", exactly, but he's more likely to say "60%" than anything else. But if there were some evidence of bias—if he were wearing a t-shirt that said, "I Hate Rain"—then I would be justified in taking $\mu_2 \neq c_1$.

The λ_is satisfy certain inequalities to ensure that c'_1 is a *bona fide* probability measure, and can be interpreted as indicators of how much independent insight v_i provides, above and beyond what was provided by $\{v_1, \ldots, v_{i-1}\}$. Namely:

$$\max\left\{\sum_{i=2}^{n} \frac{\lambda_i \mu_i}{c_1}, \sum_{i=2}^{n} \frac{\lambda_i (1-\mu_i)}{(1-c_1)}\right\} \leq 1 \tag{4.2}$$

(*N.B.* here we assume that the λ_is are positive. For the most general cases, which could include negative λ_is, there are 2^n inequalities.) The intuition is that as some λ_i gets close to 1, c_1 must be kept close to μ_i. It does not make sense for v_1 to believe, pre-disagreement, that both (1) X is very improbable, and (2) v_i, who has great insight into the truth of X, likely believes that X is very probable.

To illustrate equation (4.1) in action, let us consider the horse racing case, with $c_1 = c_2 = c_3 = 0.80$, as above. We have three epistemic peers who report the same confidence in the proposition. Here, v_1 should certainly take $\mu_2 = c_1 = 0.8$ (they are copies). μ_3, in contrast, will be somewhat less than this—say, 0.4. Of course, the distribution of c_3, in v_1's eyes, will depend on many things. But surely its mean will be less than v_1's 0.80, which is extraordinarily high in the context of horse racing. (A horse that goes off at 0.3 to win is considered a heavy favorite.)

v_1 might then evaluate $\lambda_2 = 0.01$ and $\lambda_3 = 0.30$, representing little possible epistemic help from v_2 and significant possible help from v_3. These values satisfy the necessary inequalities, and they yield a disagreement norm under which (1) v_2 exerts no epistemic influence over v_1, and, in contrast, (2) v_3's judgment *does* provide reason for v_1 to become more confident in her judgment—even though, I stress, v_1 and v_3 do not disagree about the proposition at dispute. In particular, for the values given, v_1's confidence rises from 0.80 to 0.92.

This model also prevents the bad result of equation (3.2). Again, Upco with dependence fails for $c_1 = 0.4$ and $c_2 = 0.7$. But here, with $\mu_2 = c_1 = 0.4$, we get:

$$c'_1 = 0.4 + \lambda_2(0.3) \tag{4.3}$$

One may see that the troublesome inequality, $c'_1 > c_2$, would arise if $\lambda_2 > 1$. But consistency conditions require that

$$\max\left\{\frac{0.4}{0.4-1}, \frac{0.4-1}{0.4}\right\} \leq \lambda_2 \leq \min\left\{\frac{0.4}{0.4}, \frac{1-0.4}{1-0.4}\right\} \tag{4.4}$$

The latter inequality ensures that $c'_1 \leq c_2$.

Notice how this approach satisfies our three desiderata. First, multiple epistemic interlocutors are generally admissible, and the oft-considered scenarios in which $n = 2$ are simply dealt with as special cases.

Second, we incorporate differences in competence *via* the $\lambda_i s$. To take the extreme cases, suppose v_1 believes that v_2 is epistemically infallible. Then v_1 will set $\lambda_2 = 1$, because v_2's judgment is perfectly correlated with the truth. Then $c'_1 = c_1 + (c_2 - \mu_2)$. Since $c_i, \mu_i \in (0, 1)$,[15] consistency conditions require that $c_1 = \mu_2$. Therefore, $c'_1 = c_2$. v_1 entirely abrogates her judgment and adopts v_2's opinion, as one would expect.

Next, suppose that v_1 believes that v_2 is epistemically useless; v_2's judgment is utterly uncorrelated with the truth. Then $\lambda_2 = 0$ and $c'_1 = c_1$. v_1 maintains her judgment in the face of disagreement from this interlocutor. And less extreme cases of epistemic superiority and inferiority are dealt with accordingly.[16]

The $\lambda_i s$ also allow us to incorporate the third desideratum: the possibility of dependence. We have already seen examples of this, but I would like to point out here two possible sources of dependence and how each gets handled.

First, there may be dependence between v_1 and her interlocutor(s). One imagines two weathermen (§3), each of whom makes a prediction about rain on the basis of, and only on the basis of, weather data which they both possess. Recall the interpretation of λ_i as how much independent insight v_i provides, above and beyond what was provided by $\{v_1, \ldots, v_{i-1}\}$. In this case, even if v_2 is v_1's epistemic superior (being better when it comes to

---

[15] Note that Genest and Schervish derive equation (4.1) under the assumption that the support of the c_is is [0, 1]. However, the posterior is the same in either case.

[16] West and Crosse (1992) provide a useful discussion of how to select the λ_is.

"general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias"—§1), v_2 provides *no* independent insight, and so λ_2 is set equal to 0, and so $c'_1 = c_1$.[17]

Second, there may be dependence between interlocutors. Suppose, as a variant on the horse racing case, that it is v_2 and v_3, not v_1 and v_2, who use the same handicapping methodology. Then λ_2 will be positive, because it is useful for v_1 to know that *one of* v_2 and v_3 agree with her about the winner. But it is not useful (given that knowledge) to know that *both* v_2 and v_3 agree with her. So λ_3 will be set to 0. v_1 will modify her belief only in light of one interlocutor's opinion.

5. Conclusion

Since the beginning of the disagreement debate, epistemologists have presented a number of real-world cases of disagreement which yield, variously, conciliatory and steadfast intuitions. The typical reaction has been to endorse one set of intuitions over the other and then force a theoretical structure, conforming to that set, onto the other scenarios.

For a Bayesian like me, this is misguided. When a person faces disagreement from others—whether they be peers, superiors, inferiors, or some combination thereof—she should specify likelihood functions that, in her best judgment, accurately model the circumstance she finds herself in. There is no "one size fits all" disagreement norm,

[17] This is an idealized example. We are assuming that there is no possibility of, say, misreading the radar data. If there were, $\lambda_2 > 0$ would be appropriate. We are also ignoring that the move from raw radar data to weather prediction requires judgment and experience—and thus v_2's assent carries useful information to v_1. Generally, any time that v_2 can serve as a check on v_1's work, v_1 will wish to incorporate v_2's judgment to some degree.

because each real-world case of disagreement displays different features: competence, confidence, bias, dependence, and all the rest.

I am, therefore, convinced that a Bayesian modelling approach to the epistemological problem of disagreement, as expressed in equations (2.2) (for the single interlocutor case) and (2.6) (for the multiple interlocutor case) is the only promising one.

As we have seen, this approach incorporates features which are not only alluring but apparently necessary. The toy examples in the literature, involving disagreement with a single epistemic peer, do not get at what is the real import of disagreement research: Helping human beings, in all their diversity, work together and overcome—indeed harness—differences of opinion. Under the approach I recommend, we are no longer restricted to disagreement with a single person, nor to disagreement with epistemic peers. The ubiquitous fact of dependence between judgments is handled. Bias may be explicitly modelled. And so on.

Indeed, even further generality can be incorporated. If, for example, one's epistemic interlocutors specify not just a single confidence but an entire distribution over (0, 1), that can be handled as well.[18]

And, as we have seen, steadfast and conciliatory disagreement norms fall out naturally as special cases of equations (2.2) and (2.6). When a real-world scenario evokes a conciliatory intuition, our model can provide a "conciliatory" belief revision function; and when a scenario evokes a steadfast intuition, it can provide a "steadfast" function. To make this plain, let us apply the model of §4 to two prominent scenarios from the epistemological

[18] See Clemen and Winkler 1999, Genest and Zidek 1986, and Winkler 1981.

literature, one of which is supposed to support conciliationism, and the other, steadfastness.

First, consider “Restaurant Tip”:

> Suppose that five of us go out to dinner. It’s time to pay the check, so the question we’re interested in is how much we each owe. We can all see the bill total clearly, we all agree to give a 20 percent tip, and we further agree to split the whole cost evenly, not worrying over who asked for imported water, or skipped desert [*sic*], or drank more of the wine. I do the math in my head and become highly confident that our shares are $43 each. Meanwhile, my friend does the math in her head and becomes highly confident that our shares are $45 each. How should I react, upon learning of her belief? (Christensen 2007: 193)

According to Christensen, “it seems quite clear that I should lower my confidence that my share is $43” (2007: 193). And surely that intuition—that a rational person will lose confidence in *my share is $43*—is widely shared.

Let us model this scenario. First, c_1 will be something like, say, 0.8. v_1 must round off the bill to the nearest whole number and divide that by five. These are not difficult operations for an educated adult, but it is certainly possible to make a mistake. Second, it makes sense for v_1 to take $\mu_2 \approx c_1$ (see §4). If these are five philosophers we’re talking about, they’ll be aware of the possibility of making an arithmetic error, and so they are likely to report a high, but not perfect, confidence in the proposition at issue. Third, and finally, there is the question of specifying λ_2. By the consistency conditions (and assuming non-negative correlation), $0 \leq \lambda_2 \leq 1$. v_1 is free to select the amount of epistemic weight she wishes to assign to v_2’s judgment. If v_1 wishes to treat v_2 as a pure epistemic peer, as that term is typically defined and as the case is typically interpreted, then she sets $\lambda_2 = 0.5$. This yields the following disagreement norm:

$$c'_1 \approx 0.5c_1 + 0.5c_2 \tag{5.1}$$

which is conciliationism's equal weight view—precisely the norm that has been regarded as appropriate for "Restaurant Tip".

Next, consider Jennifer Lackey's "Elementary Math":

> Harry and I, who have been colleagues for the past six years, were drinking coffee at Starbucks and trying to determine how many people from our department will be attending the upcoming APA. I, reasoning aloud, say: "Well, Mark and Mary are going on Wednesday, and Sam and Stacey are going on Thursday, and, since 2+2=4, there will be four other members of our department at that conference." In response, Harry asserts: "But 2+2 does not equal 4." Prior to this disagreement, neither Harry nor I had any reason to think that the other is evidentially or cognitively deficient in any way, and we both sincerely avowed our respective conflicting beliefs. (Lackey 2010b: 283)

The modelling is straightforward. v_1 must choose λ_2, and so she asks herself what independent epistemic insight v_2 (Harry) has into the truth of *2+2=4*. And the answer is, of course, almost none. The proposition is so simple, and so ubiquitous, and accessible to v_1 *via* so many means, that there is nothing novel that v_2 might say about it (though of course he might be unhelpful in myriad ways, if, *e.g.*, he says crazy things—as appears to be the case here). So $\lambda_2 \approx 0$, and we have the steadfast disagreement norm:

$$c'_1 \approx c_1 \tag{5.2}$$

There is the potential for flexibility here. Modify the case so that v_2 is a renowned philosopher, working on the foundations of mathematics. He is known as the smartest man who has ever lived. v_1 has the results of a recent psychiatric evaluation of v_2 which attests to his competence.

Now it is no longer obvious that v_2 has nothing useful to say about *2+2=4*. Maybe he's really discovered something profound about arithmetic. Certainly, a lesson of intellectual history is that notions long thought false, even bizarre ("time is relative") may come to be acknowledged as absolutely right. And so v_1 may take $\lambda_2 > 0$. Then, when v_2 offers his opinion, it *will* affect v_1's confidence in *2+2=4*, lowering it (if $c_2 < \mu_2$, as in "Elementary Math") or raising it (if $c_2 > \mu_2$), as appropriate.

One final point. I stress that there is no sense in which the model of equations (2.7-8), or (4.1), or any other for that matter, is correct *simpliciter*. Rather, we should give careful thought to the epistemic features of any given circumstance of disagreement (features like dependence and bias) and then choose an appropriate Bayesian model. This is the only principled way to deal with the epistemological problem of disagreement, which will otherwise, I fear, remain a perplexing one.[19]

6. References.

Arrow, K. 1950. 'A difficulty in the concept of social welfare.' *Journal of Political Economy*, 58: 328-46.

Barnett, Z. Forthcoming. 'Belief dependence: How do the numbers count?' *Philosophical Studies*.

Bergmann, M. 2009. 'Rational disagreement after full disclosure.' *Episteme*, 6: 336-53.

Christensen, D. 2007. 'Epistemology of disagreement: the good news.' *Philosophical Review*, 119: 187-217.

-----. 2009. 'Disagreement as evidence: the epistemology of controversy.' *Philosophy Compass*, 4: 756-67.

Clemen, R. T. and Winkler, R. L. 1987. 'Calibrating and combining precipitation probability forecasts.' In R. Viertl (ed), *Probability and Bayesian Statistics*, pp. 97-110. New York: Plenum.

[19] I thank the following people for their help with the matters discussed herein: David Faraci, Jesse Hill, Peter Morris, Kirun Sankaran, Jeremy Shipley, Hui Wang, and Bob Winkler. I am also grateful to audiences at the annual meetings of the Central States Philosophical Association and the Society for Exact Philosophy, and for the many helpful comments provided by anonymous referees.

-----. 1990. 'Unanimity and compromise among probability forecasters.' *Management Science*, 36: 767-79.
-----. 1999. 'Combining probability distributions from experts in risk analysis.' *Risk Analysis*, 19: 187-203.
Condorcet. 1785. 'Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.' Paris: Imprimerie Royale.
Dietrich, F. 2010. 'Bayesian group belief.' *Social Choice and Welfare*, 35: 595-626.
Easwaran, K., Fenton-Glynn, L., Hitchcock, C., and 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: 478-502.
-----. 2013. 'The puzzle of the unmarked clock and the new rational reflection principle.' *Philosophical Studies*, 164: 127-39.
Everett, T. J. 2015. 'Peer disagreement and two principles of rational belief.' *Australasian Journal of Philosophy*, 93: 273-86.
Feldman, R. 2007. 'Reasonable religious disagreements.' In L. M. Antony (ed), *Philosophers Without Gods: Meditations on Atheism and the Secular Life*, pp. 194–214. New York: Oxford University Press.
-----. 2009. 'Evidentialism, higher-order evidence, and disagreement.' *Episteme*, 6: 294-312.
French, S. 1980. 'Updating of belief in the light of someone else's opinion.' *Journal of the Royal Statistical Society*, A, 143: 43-48.
-----. 1981. 'Consensus of opinion.' *European Journal of Operational Research*, 7: 332-40.
-----. 1985. 'Group consensus probability distributions: a critical survey.' In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds), *Bayesian Statistics 2*, pp. 183-201. North-Holland: Elsevier Science.
Gardiner, G. 2014. 'The commutativity of evidence: a problem for conciliatory views of peer disagreement.' *Episteme*, 11: 83-95.
Gelfert, A. 2011. 'Who is an epistemic peer?' *Logos & Episteme*, 2: 507-14.
Genest, C. and Schervish, M. J. 1985. 'Modeling expert judgments for Bayesian updating.' *Annals of Statistics*, 13: 1198-1212.
Genest, C. and Zidek, J. V. 1986. 'Combining probability distributions: a critique and an annotated bibliography.' *Statistical Science*, 1: 114-48.
Gutting, G. 1982. *Religious Belief and Religious Skepticism*. Notre Dame: University of Notre Dame Press.
Isaacs, Y. 2019. 'A patchwork epistemology of disagreement.' *Philosophical Studies*, 176: 1873-85.
Jehle, D. and Fitelson, B. 2009. 'What is the "Equal Weight View"?' *Episteme*, 6: 280-93.
Kelly, T. 2005. 'The epistemic significance of disagreement.' In T. S. Gendler and J. Hawthorne (eds), *Oxford Studies in Epistemology, Volume 1*, pp. 167-96. Oxford: Oxford University Press.
-----. 2010. 'Peer disagreement and higher-order evidence.' In R. Feldman and T. A. Warfield (eds), *Disagreement*, pp. 111-74. Oxford: Oxford University Press.
Lasonen-Aarnio, M. 2013. 'Disagreement and evidential attenuation.' *Noûs*, 47: 767-94.
Lackey, J. 2010a. 'A justificationist view of disagreement's epistemic significance.' In A.

Haddock, A. Millar, and D. Pritchard (eds), *Social Epistemology*, pp. 298-325. Oxford: Oxford University Press.

-----. 2010b. 'What should we do when we disagree?' In T. S. Gendler and J. Hawthorne (eds), *Oxford Studies in Epistemology, Volume 3*, pp. 274-93. New York: Oxford University Press.

Lehrer, K. and Wagner, C. 1981. *Rational Consensus in Science and Society: A Philosophical and Mathematical Study*. Dordrecht: D. Reidel.

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

Lindley, D. V. 1985. 'Reconciliation of discrete probability distributions.' In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds), *Bayesian Statistics 2*, pp. 375-90. North-Holland: Elsevier Science.

Loewer, B. and Laddaga, R. 1985. 'Destroying the consensus.' *Synthese*, 62: 79-95.

Matheson, J. 2015. 'Disagreement and epistemic peers.' *Oxford Handbooks Online*, URL = <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-13>, retrieved 11 August 2018.

Morris, P. A. 1974. 'Decision analysis expert use.' *Management Science*, 20: 1233-41.

-----. 1983. 'An axiomatic approach to expert resolution.' *Management Science*, 29: 24-32.

Mulligan, T. 2015. 'Disagreement, peerhood, and three paradoxes of conciliationism.' *Synthese*, 192: 67-78.

Rawls, J. 1993. *Political Liberalism*. New York: Columbia University Press.

Shogenji, T. Manuscript. 'A conundrum in Bayesian epistemology of disagreement.'

Stone, M. 1961. 'The opinion pool.' *Annals of Mathematical Statistics*, 32: 1339-42.

Van Inwagen, P. 1996. 'It is wrong, everywhere, always, for anyone, to believe anything upon insufficient evidence.' In J. Jordan and D. Howard-Snyder (eds), *Faith, Freedom and Rationality*, pp. 137-54. Savage, MD: Rowman & Littlefield.

-----. 2010. 'We're right. They're wrong.' In R. Feldman and T. A. Warfield (eds), *Disagreement*, pp. 10-28. New York: Oxford University Press.

West, M. and Crosse, J. 1992. 'Modelling probabilistic agent opinion.' *Journal of the Royal Statistical Society*, B, 54: 285-99.

Winkler, R. L. 1968. 'The consensus of subjective probability distributions.' *Management Science*, 15: 61-75.

-----. 1981. 'Combining probability distributions from dependent information sources.' *Management Science*, 27: 479-88.

Zagzebski, L. T. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. New York: Oxford University Press.