

Disagreements, Philosophical and Otherwise

Brian Weatherson

2013

The Equal Weight View of disagreement says that if an agent sees that an epistemic peer disagrees with her about p , the agent should change her credence in p to half way between her initial credence, and the peer's credence. But it is hard to believe the Equal Weight View for a surprising reason; not everyone believes it. And that means that if one did believe it, one would be required to lower one's belief in it in light of this peer disagreement. Brian Weatherson explores the options for how a proponent of the Equal Weight View might respond to this difficulty, and how this challenge fits into broader arguments against the Equal Weight View.

This paper started life as a short note I wrote around New Year 2007 while in Minneapolis. It was originally intended as a blog post. That might explain, if not altogether excuse, the flippant tone in places. But it got a little long for a post, so I made it into the format of a paper and posted it to my website. The paper has received a lot of attention, so it seems like it will be helpful to see it in print. Since a number of people have responded to the argument as stated, I've decided to just reprint the article warts and all, and make a few comments at the end about how I see its argument in the context of the subsequent debate.

Disagreeing about Disagreement (2007)

I argue with my friends a lot. That is, I offer them reasons to believe all sorts of philosophical conclusions. Sadly, despite the quality of my arguments, and despite their apparent intelligence, they don't always agree. They keep insisting on principles in the face of my wittier and wittier counterexamples, and they keep offering their own dull alleged counterexamples to my clever principles. What is a philosopher to do in these circumstances? (And I don't mean get better friends.)

One popular answer these days is that I should, to some extent, defer to my friends. If I look at a batch of reasons and conclude p , and my equally talented friend reaches an incompatible conclusion q , I should revise my opinion so I'm now undecided between p and q . I should, in the preferred lingo, assign equal weight to my view as to theirs. This is despite the fact that I've looked at their reasons for concluding q and found them wanting. If I hadn't, I would have already concluded q . The mere fact that a friend

(from now on I'll leave off the qualifier 'equally talented and informed', since all my friends satisfy that) reaches a contrary opinion should be reason to move me. Such a position is defended by Richard Feldman (2005, 2006), David Christensen (2007) and Adam Elga (2007).

This equal weight view, hereafter EW, is itself a philosophical position. And while some of my friends believe it, some of my friends do not. (Nor, I should add for your benefit, do I.) This raises an odd little dilemma. If EW is correct, then the fact that my friends disagree about it means that I shouldn't be particularly confident that it is true, since EW says that I shouldn't be too confident about any position on which my friends disagree. But, as I'll argue below, to consistently implement EW, I have to be maximally confident that it is true. So to accept EW, I have to inconsistently both be very confident that it is true and not very confident that it is true. This seems like a problem, and a reason to not accept EW. We can state this argument formally as follows, using the notion of a peer and an expert. Some people are peers if they are equally philosophically talented and informed as each other, and one is more expert than another if they are more informed and talented than the other.

1. There are peers who disagree about EW, and there is no one who is an expert relative to them who endorses EW.
2. If 1 is true, then according to EW, my credence in EW should be less than 1.
3. If my credence in EW is less than 1, then the advice that EW offers in a wide range of cases is incoherent.
4. So, the advice EW offers in a wide range of cases is incoherent.

The first three sections of this paper will be used to defend the first three premises. The final section will look at the philosophical consequences of the conclusion.

1 Peers and EW

Thomas Kelly (2005) has argued against EW and in favour of the view that a peer with the irrational view should defer to a peer with the rational view. Elga helpfully dubs this the 'right reasons' view. Ralph Wedgwood (2007 Ch. 11) has argued against EW and in favour of the view that one should have a modest 'egocentric bias', i.e. a bias towards one's own beliefs. On the other hand, as mentioned above, Elga, Christensen and Feldman endorse versions of EW. So it certainly looks like there are very talented and informed philosophers on either side of this debate.

Now I suppose that if we were taking EW completely seriously, we would at this stage of the investigation look very closely at whether these five really are epistemic peers. We could pull out their grad school transcripts, look at the citation rates for their papers, get reference letters from expert colleagues, maybe bring one or two of them in for job-style interviews, and so on. But this all seems somewhat inappropriate for a scholarly journal. Not to mention a little tactless.¹ So I'll just stipulate that they seem to be peers in the sense relevant for EW, and address one worry a reader may have about my argument.

¹Though if EW is correct, shouldn't the scholarly journals be full of just this information?

An objector might say, “Sure it seems antecedently that Kelly and Wedgwood are the peers of the folks who endorse EW. But take a look at the arguments for EW that have been offered. They look like good arguments, don’t they? Doesn’t the fact that Kelly and Wedgwood don’t accept these arguments mean that, however talented they might be in general, they obviously have a blind spot when it comes to the epistemology of disagreement? If so, we shouldn’t treat them as experts on this question.” There is something right about this. People can be experts in one area, or even many areas, while their opinions are systematically wrong in another. But the objector’s line is unavailable to defenders of EW.

Indeed, these defenders have been quick to distance themselves from the objector. Here, for instance, is Elga’s formulation of the EW view, a formulation we’ll return to below.

Your probability in a given disputed claim should equal your prior conditional probability in that claim. Prior to what? Prior to your thinking through the claim, and finding out what your advisor thinks of it. Conditional on what? On whatever you have learned about the circumstances of how you and your advisor have evaluated the claim. (Elga 2007, 490)

The fact that Kelly and Wedgwood come to different conclusions can’t be enough reason to declare that they are not peers. As Elga stresses, what matters is the prior judgment of their acuity. And Elga is right to stress this. If we declared anyone who doesn’t accept reasoning that we find compelling not a peer, then the EW view will be trivial. After all, the EW view only gets its force from cases as described in the introduction, where our friends reject reasoning we accept, and accept reasons we reject. If that makes them not a peer, the EW view never applies. So we can’t argue that anyone who rejects EW is thereby less of an expert in the relevant sense than someone who accepts it, merely in virtue of their rejection of EW. So it seems we should accept premise 1.

2 Circumstances of Evaluation

Elga worries about the following kind of case. Let p be that the sum of a certain series of numbers, all of them integers, is 50. Let q be that the sum of those numbers is 400ϵ . My friend and I both add the numbers, and I conclude p while he concludes q . It seems that there is no reason to defer to my friend. I know, after all, that he has made some kind of mistake. The response, say defenders of EW, is that deference is context-sensitive. If I know, for example, that my friend is drunk, then I shouldn’t defer to him. More generally, as Elga puts it, how much I should defer should depend on what I know about the circumstances.

Now this is relevant because one of the relevant circumstances might be that my friend has come to a view that I regard as insane. That’s what happens in the case of the sums. Since my prior probability that my friend is right given that he has an insane seeming view is very low, my posterior probability that my friend is right should also, according to Elga, be low. Could we say that, although antecedently we regard

Wedgwood and Kelly as peers of those they disagree with, that the circumstance of their disagreement is such that we should disregard their views?

It is hard to see how this would be defensible. It is true that a proponent of EW will regard Kelly and Wedgwood as wrong. But we can't say that we should disregard the views of all those we regard as mistaken. That leads to trivialising EW, for reasons given above. The claim has to be that their views are so outrageous, that we wouldn't defer to anyone with views that outrageous. And this seems highly implausible. But that's the only reason that premise 2 could be false. So we should accept premise 2.

3 A Story about Disagreement

The tricky part of the argument is proving premise 3. To do this, I'll use a story involving four friends, Apollo, Telemachus, Adam and Tom. The day before our story takes place, Adam has convinced Apollo that he should believe EW, and organise his life around it. Now Apollo and Telemachus are on their way to Fenway Park to watch the Red Sox play the Indians. There have been rumours flying around all day about whether the Red Sox injured star player, David Ortiz, will be healthy enough to play. Apollo and Telemachus have heard all the competing reports, and are comparing their credences that Ortiz will play. (Call the proposition that he will play p .) Apollo's credence in p is 0.7, and Telemachus's is 0.3. In fact, 0.7 is the rational credence in p given their shared evidence, and Apollo truly believes that it is.² And, as it turns out, the Red Sox have decided but not announced that Ortiz will play, so p is true.

Despite these facts, Apollo lowers his credence in p . In accord with his newfound belief in EW, he changes his credence in p to 0.5. Apollo is sure, after all, that when it comes to baseball Telemachus is an epistemic peer. At this point Tom arrives, and with a slight disregard for the important baseball game at hand, starts trying to convince them of the right reasons view on disagreement. Apollo is not convinced, but Telemachus thinks it sounds right. As he puts it, the view merely says that the rational person believes what the rational person believes. And who could disagree with that?

Apollo is not convinced, and starts telling them the virtues of EW. But a little way in, Tom cuts him off with a question. "How probable," he asks Apollo, "does something have to be before you'll assert it?"

Apollo says that it has to be fairly probable, though just what the threshold is depends on just what issues are at stake. But he agrees that it has to be fairly high, well above 0.5 at least.

"Well," says Tom, "in that case you shouldn't be defending EW in public. Because you think that Telemachus and I are the epistemic peers of you and Adam. And we think EW is false. So even by EW's own lights, the probability you assign to EW should be 0.5. And that's not a high enough probability to assert it." Tom's speech requires

²This is obviously somewhat of an idealisation, since there won't usually be a unique precise rational response to the evidence. But I don't think this idealisation hurts the argument to follow. I should note that the evidence here *excludes* their statements of their credences, so I really mean the evidence that they brought to bear on the debate over whether p .

that Apollo regard he and Telemachus as Apollo's epistemic peers with regard to this question. By premises 1 and 2, Apollo should do this, and we'll assume that he does.

So Apollo agrees with all this, and agrees that he shouldn't assert EW any more. But he still plans to use it, i.e. to have a credence in p of 0.5 rather than 0.7. But now Telemachus and Tom press on him the following analogy.

Imagine that there were two competing experts, each of whom gave differing views about the probability of q . One of the experts, call her Emma, said that the probability of q , given the evidence, is 0.5. The other expert, call her Rae, said that the probability of q , given the evidence, is 0.7. Assuming that Apollo has the same evidence as the experts, but he regards the experts as experts at evaluating evidence, what should his credence in q be? It seems plausible that it should be a weighted average of what Emma says and what Rae says. In particular, it should be 0.5 only if Apollo is maximally confident that Emma is the expert to trust, and not at all confident that Rae is the expert to trust.

The situation is parallel to the one Apollo actually faces. EW says that his credence in p should be 0.5. The right reason view says that his credence in p should be 0.7. Apollo is aware of both of these facts. So his credence in p should be 0.5 iff he is certain that EW is the theory to trust, just as his credence in q should be 0.5 iff he is certain that Emma is the expert to trust. Indeed, a credence of 0.5 in p is incoherent unless Apollo is certain EW is the theory to trust. But Apollo is not at all certain of this. His credence in EW, as is required by EW itself, is 0.5. So as long as Apollo keeps his credence in p at 0.5, he is being incoherent. But EW says to keep his credence in p at 0.5. So EW advises him to be incoherent. That is, EW offers incoherent advice. We can state this more carefully in an argument.

5. EW says that Apollo's credence in p should be 0.5.
6. If 5, then EW offers incoherent advice unless it also says that Apollo's credence in EW should be 1.
7. EW says that Apollo's credence in EW should be 0.5.
8. So, EW offers incoherent advice.

Since Apollo's case is easily generalisable, we can infer that in a large number of cases, EW offers advice that is incoherent. Line 7 in this argument is hard to assail given premises 1 and 2 of the master argument. But I can imagine objections to each of the other lines.

Objection: Line 6 is false. Apollo can coherently have one credence in p while being unsure about whether it is the rational credence to have. In particular, he can coherently have his credence in p be 0.5, while he is unsure whether his credence in p should be 0.5 or 0.7. In general there is no requirement for agents who are not omniscient to have their credences match their judgments of what their credences should be.

Replies: I have two replies to this, the first dialectical and the second substantive.

The dialectical reply is that if the objector's position on coherence is accepted, then a lot of the motivation for EW fades away. A core idea behind EW is that Apollo was unsure before the conversation started whether he or Telemachus would have the most rational reaction to the evidence, and hearing what each of them say does not provide

him with more evidence. (See the ‘bootstrapping’ argument in Elga (2007) for a more formal statement of this idea.) So Apollo should have equal credence in the rationality of his judgment and of Telemachus’s judgment.

But if the objector is correct, Apollo can do that without changing his view on EW one bit. He can, indeed should, have his credence in p be 0.7, while being uncertain whether his credence in p should be 0.7 (as he thinks) or 0.3 (as Telemachus thinks). Without some principle connecting what Apollo should think about what he should think to what Apollo should think, it is hard to see why this is not the uniquely rational reaction to Apollo’s circumstances. In other words, if this is an objection to my argument against EW, it is just as good an objection to a core argument for EW.

The substantive argument is that the objector’s position requires violating some very weak principles concerning rationality and higher-order beliefs. The objector is right that, for instance, in order to justifiably believe that p (to degree d), one need not know, or even believe, that one is justified in believing p (to that degree). If nothing else, the anti-luminosity arguments in Williamson (2000) show that to be the case. But there are weaker principles that are more plausible, and which the objector’s position has us violate. In particular, there is the view that we can’t both be justified in believing that p (to degree d), while we know we are not justified in believing that we are justified in believing p (to that degree). In symbols, if we let Jp mean that the agent is justified in believing p , and box and diamond to be epistemic modals, we have the principle **MJ** (for Might be Justified).

MJ $Jp \rightarrow \Diamond Jp$

This seems like a much more plausible principle, since if we know we aren’t justified in believing we’re justified in believing p , it seems like we should at least suspend judgment in p . That is, we shouldn’t believe p . That is, we aren’t justified in believing p . But the objector’s position violates principle **MJ**, or at least a probabilistic version of it, as we’ll now show.

We aim to prove that the objector is committed to Apollo being justified in believing p to degree 0.5, while he knows he is not justified in believing he is justified in believing p to degree 0.5. The first part is trivial, it’s just a restatement of the objector’s view, so it is the second part that we must be concerned with.

Now, either EW is true, or it isn’t true. If it is true, then Apollo is not justified in having a greater credence in it than 0.5. But his only justification for believing p to degree 0.5 is EW. He’s only justified in believing he’s justified in believing p if he can justify his use of EW in it. But you can’t justify a premise in which your rational credence is 0.5. So Apollo isn’t justified in believing he is justified in believing p . If EW isn’t true, then Apollo isn’t even justified in believing p to degree 0.5. And he knows this, since he knows EW is his only justification for lowering his credence in p that far. So he certainly isn’t justified in believing he is justified in believing p to degree 0.5. Moreover, every premise in this argument has been a premise that Apollo knows to obtain, and he is capable of following all the reasoning. So he knows that he isn’t justified in believing he is justified in believing p to degree 0.5, as required.

The two replies I've offered to the objector complement one another. If someone accepts **MJ**, then they'll regard the objector's position as incoherent, since we've just shown that **MJ** is inconsistent with that position. If, on the other hand, someone rejects **MJ** and everything like it, then they have little reason to accept EW in the first place. They should just accept that Apollo's credence in *p* should be, as per hypothesis the evidence suggests, 0.7. The fact that an epistemic peer disagrees, in the face of the same evidence, might give Apollo reason to doubt that this is in fact that uniquely rational response to the evidence. But, unless we accept a principle like **MJ**, that's consistent with Apollo retaining the rational response to the evidence, namely a credence of 0.7 in *p*. So it is hard to see how someone could accept the objector's argument, while also being motivated to accept EW. In any case, I think **MJ** is plausible enough on its own to undermine the objector's position.³

Objection: Line 5 is false. Once we've seen that the credence of EW is 0.5, then Apollo's credence in first-order claims such as *p* should, as the analogy with *q* suggests, be a weighted average of what EW says it should be, and what the right reason view says it should be. So, even by EW's own lights, Apollo's credence in *p* should be 0.6.

Replies: Again I have a dialectical reply, and a substantive reply.

The dialectical reply is that once we make this move, we really have very little motivation to accept EW. There is, I'll grant, some intuitive plausibility to the view that when faced with a disagreeing peer, we should think the right response is half way between our competing views. But there is no intuitive plausibility whatsoever to the view that in such a situation, we should naturally move to a position three-quarters of the way between the two competing views, as this objector suggests. Much of the argument for EW, especially in Christensen, turns on intuitions about cases, and the objector would have us give all of that up. Without those intuitions, however, EW falls in a heap.

The substantive reply is that the idea behind the objection can't be coherently sustained. The idea is that we should first apply EW to philosophical questions to work out the probability of different theories of disagreement, and then apply those probabilities to first-order disagreements. The hope is that in doing so we'll reach a stable point at which EW can be coherently applied. But there is no such stable point. Consider the following series of questions.

Q1 Is EW true?

Two participants say yes, two say no. We have a dispute, leading to our next question.

Q2 What is the right reaction to the disagreement over Q1?

EW answers this by saying our credence in EW should be 0.5. But that's not what the right reason proponents say. They don't believe EW, so they have no reason to move their credence in EW away from 0. So we have another dispute, and we can ask

Q3 What is the right reaction to the disagreement over Q2?

³Added in 2010: I still think there's a dilemma here for EW, but I'm less convinced than I used to be that **MJ** is correct.

EW presumably says that we should again split the difference. Our credence in EW might now be 0.25, half-way between the 0.5 it was after considering Q2, and what the right reasons folks say. But, again, those who don't buy EW will disagree, and won't be moved to adjust their credence in EW. So again there's a dispute, and again we can ask

Q4 What is the right reaction to the disagreement over Q3?

This could go on for a while. The only 'stable point' in the sequence is when we assign a credence of 0 to EW. That's to say, the only way to coherently defend the idea behind the objection is to assign credence 0 to EW. But that's to give up on EW. As with the previous objection, we can't hold on to EW and object to the argument.

4 Summing Up

The story I've told here is a little idealised, but otherwise common enough. We often have disagreements both about first-order questions, and about how to resolve this disagreement. In these cases, there is no coherent way to assign equal weight to all prima facie rational views both about the first order question and the second order, epistemological, question. The only way to coherently apply EW to all first order questions is to put our foot down, and say that despite the apparent intelligence of our philosophical interlocutors, we're not letting them dim our credence in EW. But if we are prepared to put our foot down here, why not about some first-order question or other? It certainly isn't because we have more reason to believe an epistemological theory like EW than we have to believe first order theories about which there is substantive disagreement. So perhaps we should hold on to those theories, and let go of EW.

Afterthoughts

I now think that the kind of argument I presented in the 2007 paper is not really an argument against EW as such, but an argument against one possible motivation for EW. I also think that alternate motivations for EW are no good, so I still think it is an important argument. But I think it's role in the dialectic is a little more complicated than I appreciated back then.

Much of my thinking about disagreement problems revolves around the following table. The idea behind the table, and much of the related argument, is due to Thomas Kelly (2010). In the table, *S* and *T* antecedently had good reasons to take themselves to be epistemic peers, and they know that their judgments about *p* are both based on *E*. In fact, *E* is excellent evidence for *p*, but only *S* judges that *p*; *T* judges that $\neg p$. Now let's look at what seems to be the available evidence for and against *p*.

Evidence for <i>p</i>	Evidence against <i>p</i>
<i>S</i> 's judgment that <i>p</i>	<i>T</i> 's judgment that $\neg p$
<i>E</i>	

Now that doesn't look to me like a table where the evidence is equally balanced for and against p . Even granting that the judgments are evidence over and above E , and granting that how much weight we should give to judgments should track our *ex ante* judgments of their reliability rather than our *ex post* judgments of their reliability, both of which strike me as false but necessary premises for EW, it *still* looks like there is more evidence for p than against p .⁴ There is strictly more evidence for p than against it, since E exists. If we want to conclude that S should regard p and $\neg p$ as equally well supported for someone in her circumstance, we have to show that the table is somehow wrong. I know of three possible moves the EW defender could make here.

David Christensen (2011), as I read him, says that the table is wrong because when we are representing the evidence S has, we should not include her own judgment. There's something plausible to this. Pretend for a second that T doesn't exist, so it's clearly rational for S to judge that p . It would still be wrong of S to say, "Since E is true, p . And I judged that p , so that's another reason to believe that p , because I'm smart." By hypothesis, S is smart, and that smart people judge things is reason to believe those things are true. But this doesn't work when the judgment is one's own. This is something that needs explaining in a full theory of the epistemic significance of judgment, but let's just take it as a given for now.⁵ Now the table, or at least the table as is relevant to S , looks as follows.

Evidence for p	Evidence against p
E	T 's judgment that $\neg p$

But I don't think this does enough to support EW, or really anything like it. First, it won't be true in general that the two sides of this table balance. In many cases, E is strong evidence for p , and T 's judgment won't be particularly strong evidence against p . In fact, I'd say the kind of case where E is much better evidence for p than T 's judgment is against p is the statistically normal kind. Or, at least, it is the normal kind of case modulo the assumption that S and T have the same evidence. In cases where that isn't true, learning that T thinks $\neg p$ is good evidence that T has evidence against p that you don't have, and you should adjust accordingly. But by hypothesis, S knows that isn't the case here. So I don't see why this should push us even close to taking p and $\neg p$ to be equally well supported.

The other difficulty for defending EW by this approach is that it seems to undermine the original motivations for the view. As Christensen notes, the table above is specifically for S . Here's what the table looks like for T .

⁴By *ex ante* and *ex post* I mean before and after we learn about S and T 's use of E to make a judgment about p . I think that should change how reliable we take S and T to be, and that this should matter to what use, if any, we put their judgments, but it is crucial to EW that we ignore this evidence. Or, at least, it is crucial to EW that S and T ignore this evidence.

⁵My explanation is that evidence screens any judgments made on the basis of that evidence, in the sense of screening to be described below.

Evidence for p **Evidence against p**
 S 's judgment that p
 E

It's no contest! So T should firmly believe p . But that isn't the intuition anyone gets, as far as I can tell, in any of the cases motivating EW. And the big motivation for EW comes from intuitions about cases. Once we acknowledge that these intuitions are unreliable, as we'd have to do if we were defending EW this way, we seem to lack any reason to accept EW.

The second approach to blocking the table is to say that T 's judgment is an undercutting defeater for the support E provides for p . This looks superficially promising. Having a smart person say that your evidence supports something other than you thought it did seems like it could be an undercutting defeater, since it is a reason to think the evidence supports something else, and hence doesn't support what you thought it does. And, of course, if E is undercut, then the table just has one line on it, and the two sides look equal.

But it doesn't seem like it can work in general, for a reason that Kelly (2010) makes clear. We haven't said what E is so far. Let's start with a case where E consists of the judgments of a million other very smart people that p . Then no one, not even the EW theorist, will think that T 's judgment undercuts the support E provides to p . Indeed, even if E just consists of one other person's judgment, it won't be undercut by T 's judgment. The natural thought for an EW-friendly person to have in that case is that since there are two people who think p , and one who thinks $\neg p$, then S 's credence in p should be $\frac{2}{3}$. But that's impossible if E , i.e., the third person's judgment, is undercut by T 's judgment. It's true that T 's judgment will partially *rebut* the judgments that S , and the third party, make. It will move the probability of p , at least according to EW, from 1 to $\frac{2}{3}$. But that evidence won't be in any way *undercut*.

And as Kelly points out, evidence is pretty fungible. Whatever support p gets from other people's judgments, it could get very similar support from something other than a judgment. We get roughly the same evidence for p by learning that a smart person predicts p as learning that a successful computer model predicts p . So the following argument looks sound to me.

1. When E consists of other people's judgments, the support it provides to p is not undercut by T 's judgment.
2. If the evidence provided by other people's judgments is not undercut by T 's judgment, then some non-judgmental evidence is not undercut by T 's judgment.
3. So, not all non-judgmental evidence is not undercut by T 's judgment.

So it isn't true in general that the table is wrong because E has been defeated by an undercutting defeater.

There's another problem with the defeat model in cases where the initial judgments are not full beliefs. Change the case so E provides basically no support to either p or $\neg p$. In fact, E is just irrelevant to p , and the agent's have nothing to base either a firm or a probabilistic judgment about p on. For this reason, S declines to form a judgment, but T forms a firm judgment that p . Moreover, although both S and T are peers, that's because they are both equally poor at making judgments about cases like p . Here's the table then:

Evidence for p	Evidence against p
T 's judgment that p	

Since E is irrelevant, it doesn't appear, either before or after we think about defeaters. And since T is not very competent, that's not great evidence for p . But EW says that S should 'split the difference' between her initial agnosticism, and T 's firm belief in p . I don't see how that could be justified by S 's evidence.

So that move doesn't work either, and we're left with the third option for upsetting the table. This move is, I think, the most promising of the lot. It is to say that S 's own judgment *screens off* the evidence that E provides. So the table is misleading, because it 'double counts' evidence.

The idea of screening I'm using here, at least on behalf of EW, comes from Reichenbach's *The Direction of Time*, and in particular from his work on deriving a principle that lets us infer events have a common cause. The notion was originally introduced in probabilistic terms. We say that C screens off the positive correlation between B and A if the following two conditions are met.

1. A and B are positively correlated probabilistically, i.e. $\Pr(A | B) > \Pr(A)$.
2. Given C , A and B are probabilistically independent, i.e. $\Pr(A | B \wedge C) = \Pr(A | C)$.

I'm interested in an evidential version of screening. If we have a probabilistic analysis of evidential support, the version of screening I'm going to offer here is identical to the Reichenbachian version just provided. But I want to stay neutral on whether we should think of evidence probabilistically.⁶ When I say that C screens off the evidential support that B provides to A , I mean the following. (Both these clauses, as well as the statement that C screens off B from A , are made relative to an evidential background. I'll leave that as tacit in what follows.)

1. B is evidence that A .
2. $B \wedge C$ is no better evidence that A than C is.⁷

⁶In general I'm sceptical of always treating evidence probabilistically. Some of my reasons for scepticism are in Weatherson (2007).

⁷Branden Fitelson pointed out to me that the probabilistic version entails one extra condition, namely that $\neg B \wedge C$ is no worse evidence for A than C is. But I think that extra condition is irrelevant to disagreement debates, so I'm leaving it out.

Here is one stylised example of where screening helps conceptualise things. Detective Det is trying to figure out whether suspect Sus committed a certain crime. Let A be that Sus is guilty, B be that Sus's was seen near the crime scene near the time the crime was committed, and C be that Sus was at the crime scene when the crime was committed. Then both clauses are satisfied. B is evidence for A ; that's why we look for witnesses who place the suspect near the crime scene. But given the further evidence C , then B is neither here nor there with respect to A . We're only interested in finding if Sus was near the crime scene because we want to know whether he was at the crime scene. If we know that he was there, then learning he was seen near there doesn't move the investigation along. So both clauses of the definition of screening are satisfied.

When there is screened evidence, there is the potential for double counting. It would be wrong to say that if we know $B \wedge C$ we have two pieces of evidence against Sus. Similarly, if a judgment screens off the evidence it is based on, then the table 'double counts' the evidence for p . Removing the double counting, by removing E , makes the table symmetrical. And that's just what EW needs.

So the hypothesis that judgments screen the evidence they are based on, or JSE for short, can help EW respond to the argument from this table. But JSE is vulnerable to regress arguments. I now think that the argument in 'Disagreeing about Disagreement' is a version of the regress argument against JSE. So really it's an argument against the most promising response to a particularly threatening argument against EW.

Unfortunately for EW, those regress arguments are actually quite good. To see this, let's say an agent makes a judgment on the basis of E , and let J be the proposition that that judgment was made. JSE says that E is now screened off, and the agent's evidence is just J . But with that evidence, the agent presumably makes a new judgment. Let J' be the proposition that that judgment was made. We might ask now, does J' sit alongside J as extra evidence, is it screened off by J , or does it screen off J ? The picture behind JSE, the picture that says that judgments on the basis of some evidence screen that evidence, suggest that J' should in turn screen J . But now it seems we have a regress on our hands. By the same token, J'' , the proposition concerning the new judgment made on the basis of J' , should screen off J' , and the proposition J''' about the fourth judgment made, should screen off J'' , and so on. The poor agent has no unscreened evidence left! Something has gone horribly wrong.

I think this regress is ultimately fatal for JSE. But to see this, we need to work through the possible responses that a defender of JSE could make. There are really just two moves that seem viable. One is to say that the regress does not get going, because J is better evidence than J' , and perhaps screens it. The other is to say that the regress is not vicious, because all these judgments should agree in their content. I'll end the paper by addressing these two responses.

The first way to avoid the regress is to say that there is something special about the first level. So although J screens E , it isn't the case that J' screens J . That way, the regress doesn't start. This kind of move is structurally like the move Adam Elga (2010) has recently suggested. He argues that we should adjust our views about first-order matters in (partial) deference to our peers, but we shouldn't adjust our views about the right

response to disagreement in this way.

It's hard to see what could motivate such a position, either about disagreement or about screening. It's true that we need some kind of stopping point to avoid these regresses. But the most natural stopping point is the very first level. Consider a toy example. It's common knowledge that there are two apples and two oranges in the basket, and no other fruit. (And that no apple is an orange.) Two people disagree about how many pieces of fruit there are in the basket. *A* thinks there are four, *B* thinks there are five, and both of them are equally confident. Two other people, *C* and *D*, disagree about what *A* and *B* should do in the face of this disagreement. All four people regard each other as peers. Let's say *C*'s position is the correct one (whatever that is) and *D*'s position is incorrect. Elga's position is that *A* should partially defer to *B*, but *C* should not defer to *D*. This is, intuitively, just back to front. *A* has evidence that immediately and obviously entails the correctness of her position. *C* is making a complicated judgment about a philosophical question where there are plausible and intricate arguments on each side. The position *C* is in is much more like the kind of case where experience suggests a measure of modesty and deference can lead us away from foolish errors. If anyone should be sticking to their guns here, it is *A*, not *C*.

The same thing happens when it comes to screening. Let's say that *A* has some evidence that (a) she has made some mistakes on simple sums in the past, but (b) tends to massively over-estimate the likelihood that she's made a mistake on any given puzzle. What should she do? One option, in my view the correct one, is that she should believe that there are four pieces of fruit in the basket, because that's what the evidence obviously entails. Another option is that she should be not very confident there are four pieces of fruit in the basket, because she makes mistakes on these kinds of sums. Yet another option is that she should be pretty confident (if not completely certain) that there are four pieces of fruit in the basket, because if she were not very confident about this, this would just be a manifestation of her over-estimation of her tendency to err. The 'solution' to the regress we're considering here says that the second of these three reactions is the uniquely rational reaction. The idea behind the solution is that we should respond to the evidence provided by first-order judgments, and correct that judgment for our known biases, but that we shouldn't in turn correct for the flaws in our self-correcting routine. I don't see what could motivate such a position. Either we just rationally respond to the evidence, and in this case just believe there are four pieces of fruit in the basket, or we keep correcting for errors we make in any judgment. It's true that the latter plan leads either to regress or to the kind of ratificationism we're about to critically examine. But that's not because the disjunction is false, it's because the first disjunct is true.

A more promising way to avoid the regress is suggested by some other work of Elga's, in this case a paper he co-wrote with Andy Egan (Egan and Elga 2005). Their idea, as I understand them, is that for any rational agent, any judgment they make must be such that when they add the fact that they made that judgment to their evidence (or, perhaps better given JSE, replace their evidence with the fact that they made that judgment), the rational judgment to make given the new evidence has the same content as the original

judgment. So if you're rational, and you come to believe that p is likely true, then the rational thing to believe given you've made that judgment is that p is likely true.

Note that this isn't as strong a requirement as it may first seem. The requirement is not that any time an agent makes a judgment, rationality requires that they say on reflection that it is the correct judgments. Rather, the requirement is that the only judgments rational agents make are those judgments that, on reflection, she would reflectively endorse. We can think of this as a kind of ratifiability constraint on judgment, like the ratifiability constraint on decision making that Richard Jeffrey uses to handle Newcomb cases Jeffrey (1983).

To be a little more precise, a judgment is ratifiable for agent S just in case the rational judgment for S to make conditional on her having made that judgment has the same content as the original judgment. The thought then is that we avoid the regress by saying rational agents always make ratifiable judgments. If the agent does do that, there isn't much of a problem with the regress; once she gets to the first level, she has a stable view, even once she reflects on it.

It seems to me that this assumption, that only ratifiable judgments are rational, is what drives most of the arguments in Egan and Elga's paper on self-confidence, so I don't think this is a straw-man move. Indeed, as the comparison to Jeffrey suggests, it has some motivation behind it. Nevertheless it is false. I'll first note one puzzling feature of the view, then one clearly false implication of the view.

The puzzling feature is that in some cases there may be nothing we can rationally do which is ratifiable. One way this can happen involves a slight modification of Egan and Elga's example of the directionally-challenged driver. Imagine that when I'm trying to decide whether p , for any p in a certain field, I know (a) that whatever judgment I make will usually be wrong, and (b) if I conclude my deliberations without making a judgment, then p is usually true. If we also assume JSE, then it follows there is no way for me to end deliberation. If I make a judgment, I will have to retract it because of (a). But if I think of ending deliberation, then because of (b) I'll have excellent evidence that p , and it would be irrational to ignore this evidence. (Nicholas Silins (2005) has used the idea that failing to make a judgment can be irrational in a number of places, and those arguments motivated this example.)

This is puzzling, but not obviously false. It is plausible that there are some epistemic dilemmas, where any position an agent takes is going to be irrational. (By that, I mean it is at least as plausible that there are epistemic dilemmas as that there are moral dilemmas, and I think the plausibility of moral dilemmas is reasonably high.) That a case like the one I've described in the previous paragraph is a dilemma is perhaps odd, but no reason to reject the theory.

The real problem, I think, for the ratifiability proposal is that there are cases where unrati-fiable judgments are clearly preferable to ratifiable judgments. Assume that I'm a reasonably good judge of what's likely to happen in baseball games, but I'm a little over-confident. And I know I'm over-confident. So the rational credence, given some evidence, is usually a little closer to $\frac{1}{2}$ than I admit. At risk of being arbitrarily precise, let's say that if p concerns a baseball game, and my credence in p is x , the rational credence

in p , call it y , for someone with no other information than this is given by:

$$y = x + \frac{\sin(2\pi x)}{50}$$

To give you a graphical sense of how that looks, the dark line in this graph is y , and the lighter diagonal line is $y = x$.

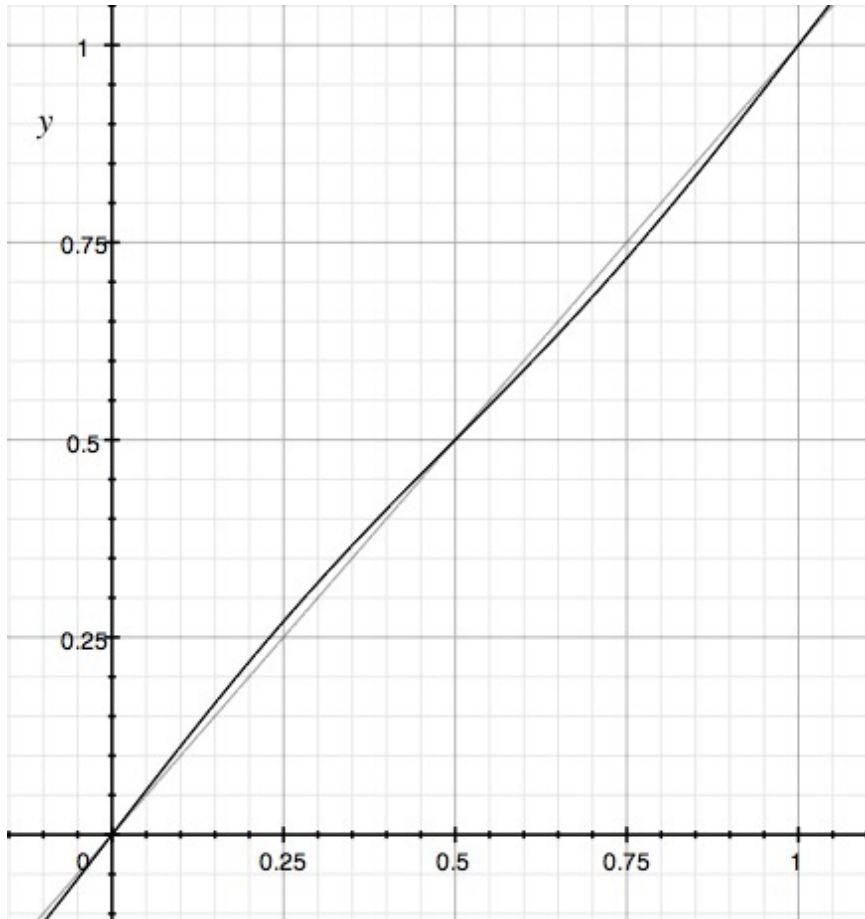


Figure 1: A sine wave around $x=y$

Note that the two lines intersect at three points: $(0, 0)$, $(\frac{1}{2}, \frac{1}{2})$ and $(1, 1)$. So if my credence in p is either 0, $\frac{1}{2}$ or 1, then my judgment is ratifiable. Otherwise, it is not. So the ratifiability constraint says that for any p about a baseball game, my credence in p should be either 0, $\frac{1}{2}$ or 1. But that's crazy. It's easy to imagine that I know (a) that in a particular game, the home team is much stronger than the away team, (b) that the stronger team usually, but far from always, wins baseball games, and (c) I'm systematically a little over-confident about my judgments about baseball games, in the way just described. In such a case, my credence that the home team will win should be

high, but less than 1. That's just what the ratificationist denies is possible.

This kind of case proves that it isn't always rational to have ratifiable credences. It would take us too far afield to discuss this in detail, but it is interesting to think about the comparison between the kind of case I just discussed, and the objections to backwards induction reasoning in decision problems that have been made by Pettit and Sugden (1989), and by Stalnaker (1996, 1998, 1999). The backwards induction reasoning they criticise is, I think, a development of the idea that decisions should be ratifiable. And the clearest examples of when that reasoning fails concern cases where there is a unique ratifiable decision, and it is guaranteed to be one of the worst possible outcomes. The example I described in the last few paragraphs has, quite intentionally, a similar structure.

The upshot of all this is that I think these regress arguments work. They aren't, I think, directly an argument against EW. What they are is an argument against the most promising way the EW theorist has for arguing that the table I started with misstates *S*'s epistemic situation. Given that the regress argument against JSE works though, I don't see any way of rescuing EW from this argument.

References

- Christensen, David. 2007. "Epistemology of Disagreement: The Good News." *Philosophical Review* 116 (2): 187–217. doi: 10.1215/00318108-2006-035.
- . 2011. "Disagreement, Question-Begging and Epistemic Self-Criticism." *Philosophers' Imprint* 11 (6): 1–22. <http://hdl.handle.net/2027/spo.3521354.0011.006>.
- Egan, Andy, and Adam Elga. 2005. "I Can't Believe I'm Stupid." *Philosophical Perspectives* 19 (1): 77–93. doi: 10.1111/j.1520-8583.2005.00054.x.
- Elga, Adam. 2007. "Reflection and Disagreement." *Noûs* 41 (3): 478–502. doi: 10.1111/j.1468-0068.2007.00656.x.
- . 2010. "How to Disagree about How to Disagree." In *Disagreement*, edited by Ted Warfield and Richard Feldman, 175–87. Oxford: Oxford University Press.
- Feldman, Richard. 2005. "Respecting the Evidence." *Philosophical Perspectives* 19 (1): 95–119. doi: 10.1111/j.1520-8583.2005.00055.x.
- . 2006. "Epistemological Puzzles about Disagreement." In *Epistemology Futures*, edited by Stephen Cade Hetherington, 216–26. Oxford: Oxford University Press.
- Jeffrey, Richard C. 1983. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.
- Kelly, Thomas. 2005. "The Epistemic Significance of Disagreement." *Oxford Studies in Epistemology* 1: 167–96.
- . 2010. "Peer Disagreement and Higher Order Evidence." In *Disagreement*, edited by Ted Warfield and Richard Feldman, 111–74. Oxford: Oxford University Press.
- Pettit, Philip, and Robert Sugden. 1989. "The Backward Induction Paradox." *Journal*

- of *Philosophy* 86 (4): 169–82. doi: 10.2307/2026960.
- Silins, Nicholas. 2005. “Deception and Evidence.” *Philosophical Perspectives* 19: 375–404. doi: 10.1111/j.1520-8583.2005.00066.x.
- Stalnaker, Robert. 1996. “Knowledge, Belief and Counterfactual Reasoning in Games.” *Economics and Philosophy* 12: 133–63. doi: 10.1017/S0266267100004132.
- . 1998. “Belief Revision in Games: Forward and Backward Induction.” *Mathematical Social Sciences* 36 (1): 31–56. doi: 10.1016/S0165-4896(98)00007-9.
- . 1999. “Extensive and Strategic Forms: Games and Models for Games.” *Research in Economics* 53 (3): 293–319. doi: 10.1006/reec.1999.0200.
- Weatherson, Brian. 2007. “The Bayesian and the Dogmatist.” *Proceedings of the Aristotelian Society* 107: 169–85. doi: 10.1111/j.1467-9264.2007.00217.x.
- Wedgwood, Ralph. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.

Published in *The Epistemology of Disagreement: New Essays*, 2013, pp. 54–77.