

Self-Exempting Conciliationism is Arbitrary



SIMON BLESSENOHL

Abstract

Self-exempting conciliationism is the view that it is rational to give weight to the opinions of peers in disagreement, except in disagreements about how to respond to disagreement. The special treatment of disagreements about disagreement, which is important to avoid self-undermining, seems arbitrary. Two arguments against this objection were put forward. Elga [3] aims to show that there is an independent motivation for conciliationism to be self-exempting. Pittard [5] argues that the special treatment is not arbitrary because the concern for epistemic deference motivates conciliatory responses only in ordinary disagreements, but not in disagreements about disagreement. I will argue that both replies fail—none of them can provide a consistent justification for why one ought to be conciliatory in disagreements except in cases of disagreement about disagreement.

Keywords: *disagreement, conciliationism, self-undermining, arbitrariness objection*

1 Self-Exempting Conciliationism, the Arbitrariness Objection, and Elga's Reply

Self-exempting conciliationism is a view on how to respond to peer disagreement. A peer is a person who has the same evidence about the proposition in question and approximately the same ability to form judgements about the sort of propositions at issue. If you find out that a peer has come to a different conclusion on the basis of the same evidence, how should you respond? Steadfast views hold that rationality does not require you to adjust your credence towards that of your peer, while conciliatory views claim that it does.

Self-exempting conciliationism claims that the following rule expresses the rational way to respond to peer disagreement: be conciliatory in cases of disagreement, but if the disagreement is about how to respond to disagreement, i.e. about the truth of conciliationism and competing views, remain steadfast. That is, your confidence in conciliationism should not be diminished in response to the fact that peers hold that conciliationism is wrong. In all other disagreements, however, you ought to move your credence towards that of your peers. Self-exempting conciliationism was proposed to overcome the problem of self-undermining which unrestricted conciliatory views seem to face. Self-undermining is unpleasant because it means that conciliationism is a view that one cannot hold at the moment since there are many able philosophers doubting the truth of conciliationism, and as a conciliationist, it seems necessary to take their opinions into account. But this, of course, does not mean that conciliationism is false—the current epistemic climate might just be unfavourable for a rational subject to believe the truth about how to respond to disagreement, namely conciliatorily. Much more problematic is that, given certain assumptions, a conciliatory rule which self-undermines gives inconsistent advice in some situations, as will be explained in more detail later.

The arbitrariness objection against self-exempting conciliationism is that “[i]t looks arbitrary for a view to recommend that one be conciliatory about most matters, but not about disagreement itself” [3, p. 184]. There has been some controversy about what ‘arbitrary’ means here. I propose to focus on what should uncontroversially be acknowledged as the core of the arbitrariness objection, namely that self-exempting conciliationism lacks a justification, even if one accepted all the arguments that were put forward for being conciliatory. More precisely, it lacks a *consistent* justification, a justification which does not make use of arguments yielding contradictory conclusions.¹ It seems to lack such a consistent justification because the arguments from conciliationists seem to imply that one ought to be conciliatory in cases of disagreement about disagreement, too. That is, while there are proposed justifications for *always* being conciliatory, there does not seem to be a good argument for the rule of always being conciliatory, except in cases of disagreement about disagreement. Crucially, the self-exempting conciliationist cannot just keep endorsing the usual arguments for being conciliatory to justify that part of the rule which recommends conciliatory responses in ordinary cases of disagreement. As these arguments also seem to apply to disagreement about disagreement, proponents of self-exempting

conciliationism must concede that there is something wrong with them since these arguments suggest, in at least one case, a response which the self-exempting conciliationist rejects as irrational: being conciliatory in disagreements about disagreement. Thus, there seem to be two ways to properly answer the arbitrariness objection. One way would be to show that the usual arguments for conciliationism, contrary to what it looks like at first glance, suggest being steadfast, not conciliatory, in cases of disagreement about disagreement, as well as being conciliatory in all other cases of disagreement. Another way would be to stop endorsing the usual arguments for conciliationism and rely on entirely different considerations for justifying self-exempting conciliationism. Both ways would yield a consistent justification for self-exempting conciliationism. It should be noted that a different approach would be to endorse arguments for always being conciliatory and arguments for being steadfast in disagreements about disagreement, and claim that disagreements about disagreement are situations in which different epistemic ideals, all of which are perfectly correct, impose contradicting requirements on us. This line of argument, pursued by Christensen [2], will not be discussed in this paper.

After having introduced self-exempting conciliationism, Elga [3] replies to the arbitrariness objection by first showing that any inductive method has to be dogmatic with respect to its own correctness because otherwise it would give logically inconsistent advice in some situations. Inductive methods are (sets of) rules that give advice on how to respond to courses of evidence. An inductive method is dogmatic about its own correctness iff there is no course of evidence so that the inductive method recommends to believe that the inductive method is incorrect (i.e. that following it is not the rational way of taking evidence into account). Rules about how to respond to disagreement are components of inductive methods, giving advice on how to respond to evidence obtained from disagreement situations. Hence, a rule about how to respond to disagreement must be dogmatic with respect to its correctness: if a part of an inductive method is not dogmatic about its own correctness, then the inductive method, seen as a whole, is not dogmatic about its own correctness because correctness of the whole inductive method entails correctness of its parts. Crucially, this requirement of being dogmatic with respect to its own correctness holds for any rule about disagreement, not just for conciliatory rules.

Then, Elga notes that for conciliatory rules to be dogmatic about their own correctness, they must recommend steadfast responses in cases

of disagreements about disagreement. If they recommended a conciliatory response, they would recommend to reduce one's confidence in themselves, because such rules are exactly what is at dispute in disagreements about disagreement. Such an unrestricted conciliatory rule would not be dogmatic about its own correctness.

Elga [3, p. 185] concludes that the self-exemption of self-exempting conciliationism is not arbitrary: any view on disagreement “has good independent motivation to avoid calling for its own rejection”. Therefore, in particular, there is a good independent motivation for a conciliatory rule to refrain from giving conciliatory advice in cases of disagreement about disagreement, and instead recommend being steadfast.

This reply has been criticized in multiple ways (see [5, 8]). From the clarification of the arbitrariness objection given above, it should be obvious that it fails to properly address what is at issue. Even if it is taken for granted that Elga's argument succeeds in justifying steadfast responses in cases of disagreement about disagreement, it still does not give a complete justification of self-exempting conciliationism since the question remains how the conciliatory responses in ordinary disagreements are justified. As noted above, the usual arguments for conciliationism seem to entail the rationality of conciliatory responses in cases of disagreement about disagreement, too, and hence are not available, without modification, to the self-exempting conciliationist. Taking Elga's argument as justification for the self-exempting part of self-exempting conciliationism and relying on the usual arguments for conciliationism in ordinary cases of disagreement would result in an inconsistent justification for self-exempting conciliationism, consisting of arguments with contradictory conclusions. Without addressing the issue of how these arguments fit together, this is not a proper reply to the arbitrariness objection.

Besides criticizing other parts of Elga's argument, Pittard [5] also takes it not to answer the arbitrariness objection. Pittard understands the objection as the charge that “*from the perspective of the concerns motivating conciliationism*, there is no reason to restrict conciliatory requirements to topics other than conciliationism” [5, p. 7]. This fits the perspective on the arbitrariness objection I suggested. If the rule is not motivated by the concerns motivating conciliationists, then the conciliatory part of self-exempting conciliationism cannot be justified by the arguments that conciliationists have put forward. However, if it could be shown that self-exemption was motivated by the concerns motivating conciliationism, then the usual arguments for conciliationism would make self-exempting conciliationism as justified as unrestricted

conciliatory rules seemed to have been before. Pittard aims to show that this is the case. The proper reply to the arbitrariness objection is that self-exempting conciliationism is not arbitrary because the motivation for conciliationism, the concern for epistemic deference, actually motivates conciliatory responses in all cases but disagreement about disagreement.

In the next section, a closer look at this argument will be taken. In the two following sections, I will argue that it fails.

2 *The Argument for Self-Exemption being Motivated by Epistemic Deference*

Suppose Alice, a conciliationist, has “a commitment to showing epistemic deference that is proportionate to the epistemic qualifications of different thinkers, so that the views of an epistemic peer are given equal weight to one’s own views” [5, p. 8]. Alice meets Bob, whom she considers to be a peer with respect to questions about peer disagreement—that is, he has the same evidence (in this case, philosophical arguments) and comparable abilities relevant to doing philosophy. Bob says that he thinks conciliationism is false. Does it follow from Alice’s commitment to epistemic deference that she should respond by becoming less confident in conciliationism?

To answer this question, Pittard suggests that when assessing how deferential an agent responds to a disagreement—no matter what the disagreement is about—, there are two things to consider (see [5, p. 9]). Firstly, one can consider the new credence that the agent decides to adopt in response to the disagreement: how much did she adjust her credence towards that of her peer? If this question is asked, one considers *deference on the credence level*. But there is another aspect that can be considered: how did the agent come to the decision to adopt this particular credence in response to the disagreement? More precisely, what is the reasoning process that the agent would have to go through to reason from the fact that her peer disagrees with her to the conclusion that she ought to adjust her credence as she did? For example, the reasoning of a conciliationist might be made explicit as follows: “I believe that I should give equal weight to my credences and those of my peers. Bob is a peer and thinks p is false. I think p is true. So I ought to adjust my credence in p to 0.5.” To show *deference on the reasoning level* means not to rely on reasoning that your peer rejects.

With these two dimensions of assessing deference in disagreement situations, Pittard evaluates the possible responses of Alice. Suppose

she reduced her credence to 0.5 by giving equal weight to her pre-disagreement credence, which is close to 1, and Bob's credence, which is close to 0. Then, on the *credence level*, she has certainly exhibited epistemic deference: she gave equal weight to her pre-disagreement credence and that of her peer. However, on the *reasoning level* she was non-deferential. This is because she relied on conciliationism being true when she reasoned about how she ought to adjust her credences. She assumed that conciliationism was true and thus came to the conclusion that she ought to reduce her credence to 0.5. Thereby, she relied on a premise which Bob rejects. While giving weight to Bob's views on the credence level, she did not give weight to his views on the reasoning level.

Now suppose Alice remained steadfast, i.e. did not reduce her confidence in conciliationism. Then, on the credence level, Alice was not deferential because she did not adjust her credence towards Bob's credence. However, she was deferential on the reasoning level because she avoided relying on the assumption that conciliationism is true, which Bob denies—her “non-reliance on a conciliatory rationale [...] is deferential” [5, p. 10].²

According to Pittard, “there is no evident reason for privileging deference at either the credence level or the reasoning level” [5, p. 9]. Thus, it is not more deferential to be deferential on the credence level and reduce the credence in conciliationism, than to be deferential on the reasoning level and keep a credence of 1. But this means that the concern for epistemic deference does not give Alice a reason to lower her credence. In situations in which Alice is confronted with a peer rejecting conciliationism, deference on the credence level trades off with deference on the reasoning level, so the commitment to epistemic deference does not give Alice a reason to become less confident in conciliationism. Therefore, as long as conciliationists are motivated by a concern for epistemic deference, it is not arbitrary for them to be steadfast in cases of disagreement about disagreement; in fact, for them, “the most reasonable conciliationism is a resolute [i.e. self-exempting] conciliationism” [5, p. 2]. For if a situation of disagreement about disagreement gives the conciliationist no reason to become less confident in conciliationism, then doing so instead of just continuing to rely on other reasons would seem irrational.³

This argument is supposed to work no matter which competing view about disagreement Bob endorses (see [5, p. 10]). In particular, it does not depend on Bob endorsing steadfastness. He could reject steadfastness, too, and favour, for example, the right reasons view, which licenses a steadfast response only to that party of the disagreement who judged

the shared pre-disagreement evidence correctly (see [4]). Even if Bob believed in the right reasons view, deference on the credence level would trade off with deference on the reasoning level: if Alice adjusted her credence, she would rely on conciliationism being true, a premise which Bob rejects. The steadfastness which Alice exhibits is a consequence of her not relying on conciliationism being true, not a consequence of her assuming that steadfast views are correct.

To sum up, the argument aims to show that a self-exempting conciliatory rule is not arbitrary, provided that it is motivated by a concern for epistemic deference. Such a concern motivates conciliatory responses only in cases of ordinary disagreements. In the special case of disagreement about disagreement, it does not motivate reducing the credence in conciliationism. To be precise, Pittard's conclusion is that self-exempting conciliationism is not arbitrary, from the point of view of conciliationism, as long as conciliationism endorses the following claim:

- (ED) It is rational to exhibit epistemic deference to peers both on the reasoning level, by not relying on disputed premises, and on the credence level, by adjusting credences towards those of the peers. None of the two levels is privileged, so if they trade off, the rationality of deference does not give a reason to favour any of the possible responses.

In the following, this particular interpretation of the commitment to epistemic deference will be referred to by (ED), to distinguish it from other possible ways of understanding epistemic deference.

In the rest of this paper, I will criticize Pittard's argument in two ways. Firstly, the argument would only lead to a consistent justification of self-exempting conciliationism if, as Pittard seems to assume, the usual arguments for the rationality of epistemic deference can be used to justify (ED). Then, self-exempting conciliationism would be justified by these arguments via (ED). I will argue that this is not the case: the way epistemic deference is construed in the suggested reply to the arbitrariness objection cannot be justified by the usual argument for deference. Thus, self-exempting conciliationism is still as unjustified as before because presently there is no argument for (ED). But then, the reply fails to address the arbitrariness objection: the price for arguing that self-exemption is justified by epistemic deference, the very same principle which justifies conciliatory responses in ordinary disagreements, is to

construe deference in a way that seems to make it unjustified. Secondly, I will argue that even if (ED) could be justified, it actually does not motivate self-exemption in all cases of disagreement about disagreement. While it can justify a steadfast response if the peer rejects conciliationism, it cannot justify keeping full credence in conciliationism if the peer has some doubts about conciliationism, but does not entirely reject it. In such cases, the rule which (ED) motivates recommends to reduce the credence in the rule itself. If this is true, (ED) can only motivate a partially self-exempting rule, and partially self-exempting conciliatory rules are inconsistent.

It should be emphasized that Pittard does not commit himself to (ED). He only argues that it is plausible that conciliationists endorse (ED), and if they do, then they have a good motivation to follow a self-exempting conciliatory rule. However, if my first objection is right, then the standard justification for conciliatory views does not justify (ED), so it seems unlikely that conciliationists would actually endorse (ED).

3 Objection 1: Justifications For Epistemic Deference do not Support (ED)

Conciliationists usually defend their views by arguing that a disagreement provides new evidence, which, though indirectly, bears on the proposition that the disagreement is about:

“The main motivation for Conciliatory positions [...] begins with two thoughts: that the peer’s disagreement gives one evidence that one has made a mistake in interpreting the original evidence, and that such evidence should diminish one’s confidence [...].” [1, p. 757]

Thus, to argue for the rationality of epistemic deference, giving weight to the views of your peers, conciliationists appeal to an evidence-based perspective on disagreement. In the following, I will argue that an evidence-based perspective does not justify (ED). I will firstly sketch what an attempted justification of (ED) from the evidence-based perspective would look like, and then criticize it.

Note that if (ED) should be justified from an evidence-based perspective, then, in particular, one implication of (ED) has to be justified: in cases of disagreement about disagreement, one does not have a reason to reduce one’s confidence in conciliationism. Justifying this from the evidence-based perspective amounts to showing that in cases like the

one of Alice and Bob, the evidence which Alice gets does not give her a reason to diminish her confidence. That is, in terms of making a rational response to this evidence, the response of diminishing her credence is not preferred to staying steadfast. In more detail, an attempted argument for (ED) from the evidence-based perspective has to show that the evidence Alice gets from the disagreement about disagreement is special: *prima facie*, diminishing her confidence (deference on the credence level) looks like a good thing in terms of responding to the evidence—in other cases of disagreement this is what the evidence gives one a reason to, as the quote above suggests. But in the case of disagreement about disagreement, the evidence is special in a way that makes avoiding to reason based on a conciliatory rationale (deference on the reasoning level), and hence not diminishing her credence, just as desirable.

Let me attempt to sketch how a proponent of (ED) might argue for this. In the case of disagreement about disagreement, Alice gets evidence which indicates that she has made a mistake when she became convinced of conciliationism on the basis of current philosophical arguments. The disagreement evidence indicates that, in fact, the original evidence (philosophical arguments) might well suggest that conciliationism is false and steadfastness is true—and Alice mistakenly thought the opposite.

Prima facie, it looks like it would be the proper response for Alice to reduce her credence in conciliationism upon receiving this piece of evidence. But, one might argue, responding to the evidence by diminishing her credence is actually bad in terms of rationally responding to evidence: after all, Alice would rely on the premise that conciliationism is true when she reasons to the conclusion that she ought to diminish her confidence in response to the disagreement. But this premise is exactly what the evidence denies! Therefore, she would not respect the evidence if she reasoned in that way. Thus, both possible responses that Alice could make seem to have downsides when looked at from the evidence-based perspective.

Up to this point, the argument from the evidence-based perspective shows how the first part of (ED), the importance of two levels of deference, comes into play in the case of disagreement about disagreement. It shows that from the evidence-based perspective, it makes sense to speak of two levels of deference because these correspond to two ways in which one can respect evidence. Deference on the credence level, adjusting your credence towards that of your peer, corresponds to increasing (or diminishing) your credence in the proposition which the evidence supports

(or speaks against). Deference on the reasoning level, not relying on premises which your peer disputes, corresponds to not using premises in the reasoning about how to respond to the evidence which the evidence itself indicates to be false. Furthermore, the argument showed that in the special case of disagreement about disagreement, the evidence-based perspective also recognizes a trade-off between the two levels of deference: Alice cannot respect the evidence in both ways at the same time.

For the rest of this section, I will assume that the argument so far succeeds. But the argument has to be continued to show the last, crucial part of (ED): in case of a trade-off between the two levels, none is privileged, and so the evidence does not give a reason for any particular of the possible responses. If one succeeded in such a completion of the argument, then one would have shown that Alice has, from the evidence-based perspective, no reason to adopt a lower credence than her current credence. But if the disagreement provides no reason for her to modify her credence, she would be well-advised, as (ED) implies, to just keep her current credence. The rest of this section is dedicated to whether it is possible to continue the argument above to justify this crucial last part of (ED) from the evidence-based perspective.

Continuing the argument, a proponent of (ED) might point out that the trade-off in fact reveals that the evidence Alice receives is, in a sense, self-defeating. That is, from the evidence-based perspective, disagreement about conciliationism is fundamentally different from ordinary disagreement because it gives self-defeating evidence. That there are competing ways of respecting the evidence in disagreements about disagreement is a manifestation of that fact.

So why is the evidence which Alice receives self-defeating? The evidence that Alice receives when she gets to know that Bob does not believe in conciliationism is evidence for the proposition that current philosophical arguments in fact support the conclusion that conciliationism is wrong, and thus it is evidence which indirectly supports the conclusion that conciliationism is wrong. But, it might be argued, if conciliationism was wrong, then evidence from disagreements would not be (reliable) evidence at all. This might be easier to see in its contrapositive form: if evidence from disagreements was proper evidence, then surely some form of conciliationism would be true. So, what the piece of evidence Alice receives says is that conciliationism is false, and, by implication, that alleged pieces of evidence obtained from disagreement, *like that very piece of evidence*, are in fact no evidence. In a generalized terminology, the piece of evidence says: “The view that pieces of

evidence like I are proper evidence is wrong". But such evidence, one might say, is self-defeating. Clearly, rationality does not require one to change one's beliefs in response to self-defeating evidence, so Bob disagreeing with Alice does not give Alice a reason to lower her credence in conciliationism.

If this completion of the argument went through, then it would seem that the crucial implication of (ED) concerning disagreement about disagreement can be justified by the evidence-based perspective on disagreement. It would be justified that the disagreement with Bob does not give Alice a reason to lower her credence.

This concludes my attempt to reconstruct an argument for (ED) from the evidence-based perspective. I started this section by pointing out that the rationality of deference is usually argued for by adopting a perspective on disagreement which considers it to provide new evidence. As it is not obvious whether (ED) is justified from this perspective, I attempted to sketch an argument from this perspective that would justify (ED), in particular its implication concerning disagreements about disagreement. I will now object to the argument I have sketched, and thereby give what I believe to be the correct conclusions to draw from an evidence-based perspective on disagreement about disagreement. In short, the problem lies in the assumption that because the piece of evidence which Alice obtains is self-defeating, she does not have to respond to it by lowering her credence in conciliationism. That is, even if one accepted the entire argument up to the point of the evidence being, in a way, self-defeating, I will argue that the kind of self-defeat involved here clearly demands of her to reduce her credence in conciliationism, and that, therefore, the evidence-based perspective in fact suggests that (ED) is incorrect.

To show this, consider the evidence-based perspective on disagreement about disagreement in a generalized version. Some mechanism X produces outputs which you believe to be reliable pieces of evidence, i.e. evidence which you believe you ought to take into account. One day, it generates evidence supporting the proposition that following the evidence which X produces is not rational, that the evidence which X produces is unreliable. Now, does rationality require you to respond to this by reducing confidence in evidence generated by X , or does the evidence not give you a reason to adjust your credence in evidence generated by X , so that you ought to continue being fully confident in past and future evidence from X ?

I think there are strong intuitions that the answer is that the new

evidence ought to diminish the confidence in evidence generated by *X*. Consider some instantiations of this general situation. First, let *X* be perception. Suppose you believe in perception giving you reliable evidence. One day you get evidence via perception that perception is less reliable than you thought. Intuitions are that it would be irrational to ignore this and keep being as confident in perception as before.⁴ Maybe one should not lose much confidence in perception by such a single piece of evidence, but not taking this perceptual evidence into account at all seems irrational. Or suppose *X* is a book which lists putative facts about the world, and you consider it to be quite reliable. One day, you find a sentence in it stating that sentences in the book are entirely fictitious, i.e. that one should not treat them like evidence. Intuitions are that this should somehow affect how much you trust what you read in the book. I am not claiming that you should completely stop trusting the book, only that rationality requires you to lose some amount of confidence in what is in the book. And, analogously, if *X* is evidence generated from disagreement situations, then disagreement about disagreement, from the evidence-based perspective, requires a reduction of the credence in conciliationism. Continuing to regard disagreement evidence as reliable as before encountering the disagreement about disagreement would be irrational.

What, then, is the flaw in the argument for (ED) given above, which seems to establish the opposite conclusion? I think that, if the argument sounded convincing in the first place, this is due to a failure to distinguish directly self-defeating evidence and indirectly self-defeating evidence. The problematic piece of evidence is not directly self-defeating in the sense that all it says is ‘This piece of evidence is misleading’. In such a case of direct self-defeat, intuitions might be that it is rational not to adjust any credences in response to it.⁵ But in the case under consideration, the evidence does not only state the unreliability of itself, but of *all* evidence generated by *X*, including, but not limited to, itself. In such cases, as I hope to have made plausible, staying as confident in evidence from *X* as before would be irrational. *X* suggesting that output from *X* is unreliable ought to have some impact on how future and past output of *X* is dealt with, and thus the credence in the proposition that *X* is reliable ought to be diminished. And this is exactly because *X* is not only saying that the current output is unreliable, but that all output from *X* is unreliable. The argument for (ED) given above fails to note the indirectness of self-defeat in the case of disagreement about disagreement. But if this fact is appreciated, then it becomes clear that

from the self-defeat in this particular situation one cannot conclude that rationality does not require adjusting any beliefs in response to the evidence.

There might be an uneasy feeling left concerning my claim that a rational response to the problematic piece of evidence from X would involve some reduction of the credence in the reliability of evidence from X : for if you reduced your credence in evidence from X , should you not, with hindsight, consider your response to that piece of evidence as irrational? Thus, just as in the special case of disagreement about disagreement, the danger of incoherence arises: incoherence between how you believe you ought to respond to evidence from X and how you actually do. However, there is a (non-zero) amount by which the credence in evidence from X can be reduced so that no incoherence arises. The argument for this would be analogous to the argument that one can coherently get less confident in conciliationism in cases of disagreement about disagreement which was given by Weatherson [7]. I have not claimed that it is rational to give the indirectly self-defeating evidence as much weight as one gave to evidence of X in the past, just that it must have *some* impact on how to respond to other outputs of X , i.e. some impact on your credence in the reliability of X —and this is possible without being incoherent.

In summary, (ED) cannot be motivated from the evidence-based perspective. More precisely, from this perspective it cannot be motivated that the trade-off between deference on the reasoning level and deference on the credence level, which occurs in disagreements about disagreement, means that one does not have a reason to show any deference on the credence level. The evidence-based perspective strongly suggests that one should get less confident in conciliationism, i.e. treat disagreements as a less reliable source of evidence than one did before. This means that (ED) cannot be motivated by the central argument for the rationality of epistemic deference.⁶

If my argument is correct, then the justification problem raised by the arbitrariness objection is not solved by the move Pittard suggests. Even if self-exempting conciliationism was shown to be motivated by the commitment to (ED), this commitment itself seems to lack a proper justification. While the common notion of epistemic deference, which only applies to the credence level, might be considered justified if one accepts the evidence-based view on disagreement situations, such a strategy does not work for (ED).

Of course, my argument, if true, only shows that the usual lines of argument for deference do not justify (ED). It does not exclude that

conciliationists come up with new arguments for (ED). However, as long as this has not happened, the justification problem of self-exempting conciliationism has not been solved because the old arguments no longer work. But even if (ED) could be justified, little would be gained because, as I will now argue, (ED) does not motivate a consistent, fully self-exempting conciliatory rule.

4 Objection 2: (ED) does not Motivate Full Self-Exemption

So far, it was assumed that Bob rejects conciliationism.⁷ Now, suppose that while Bob's credence is smaller than Alice's, he does not reject conciliationism. For instance, Bob might have a credence of 0.7 in conciliationism, while Alice's credence is close to 1. That is, Bob thinks that the arguments for conciliationism are quite good, but they leave more room for doubt than Alice concedes.

My objection is that in such a case, (ED) motivates Alice to get less confident in conciliationism. Deference on the reasoning level was understood as not relying on premises that the peer rejects. Since Bob does not reject conciliationism, it seems that it would not be non-deferential for Alice to rely on conciliationism in her reasoning. Thus, deference on the credence level does not trade off with deference on the reasoning level. Therefore, Alice's commitment to (ED) requires her to lower her confidence in conciliationism in response to her disagreement with Bob. But then, (ED) does not motivate self-exemption in all cases of disagreement about disagreement. This suffices to make the rule motivated by (ED) inconsistent.

The last step of the objection, concluding that the partially self-exempting rule which (ED) motivates is inconsistent, might not be obvious. After all, Elga has given the inconsistency argument only in a setting of absolute beliefs: he has argued that any rule which recommends abandoning itself and following a different rule is inconsistent. But Alice could never be lead to *abandon* conciliationism by responding conciliatorily to disagreements in which the peers do not reject conciliationism—and these are the only cases for which I claim that (ED) cannot justify self-exemption. She could only be made to become less confident in conciliationism, but not to a degree at which she rejects it. The idea why inconsistency occurs no matter how little the confidence in the rule is reduced, is, as Pittard [5, pp. 3–4] agrees, that if a conciliatory rule recommends to get less confident in itself, the rule would automatically advise to follow a rule which is less conciliatory than itself, and hence

issue inconsistent advice. I will now spell out this idea, giving a straightforward generalization of Elga's inconsistency argument from absolute beliefs to degrees of belief. I will illustrate the point using the example of the partially self-exempting rule motivated by (ED), but it should be obvious that an argument of this form can show that, generally, if a rule advises to reduce the confidence in itself and have some non-zero credence in a competing rule, then self-undermining occurs.

Let R be the rule motivated by (ED): always give equal weight to your credence and that of your peer, but if your peer rejects conciliationism, be steadfast. Suppose that Bob has a degree of belief of 0.7 in R and 0.3 in steadfastness, while Alice's degree of belief in conciliationism is close to 1.0 before meeting Bob.⁸ Since Bob does not reject conciliationism, R gives conciliatory advice, saying that in light of the disagreement with Bob, Alice should believe R with a confidence of $0.5 \cdot 0.7 + 0.5 \cdot 1.0 = 0.85$, and steadfastness with a confidence of $0.5 \cdot 0.3 + 0.5 \cdot 0.0 = 0.15$. That is, it advises Alice to give equal weight to her pre-disagreement degrees of belief and Bob's degrees of belief, both with respect to the proposition that conciliationism is true and that steadfastness is true. Hence, after meeting Bob, Alice has (small) doubts about the correctness of R , and assigns some non-zero credence, 0.15, to a competing view, steadfastness. But if R recommends these credences in R and steadfastness, then it recommends having credences which are a mixture of the responses of R and steadfastness, weighted by 0.85 and 0.15, respectively. It is assumed here that if there is uncertainty about which rule to follow, the reliance on them ought to be proportioned according to the confidence in them. Elga uses such an assumption when saying that rules give *indirect* advice in this fashion, Pittard makes the same assumption by saying that "a *reasonable* inductive method will require that my reliance on competing methods be proportioned to my credences for those methods" [5, p. 3]. For someone who rejects such assumptions, the need for self-exemption does not even arise because unrestricted conciliationism would not issue inconsistent advice if it did not give indirect advice in this way. But by only relying on the same assumptions as Pittard, it follows that R , the rule motivated by (ED), (indirectly) recommends to believe R to a degree of $0.85 \cdot 0.85 + 0.15 \cdot 1.0 \approx 0.92$. This is because it indirectly recommends to adjust the credence in R according to a mixture of the responses of the different, competing rules which Alice has non-zero credence in, i.e. R and a steadfast rule. The response recommended by R , adopting a credence of 0.85 in R , and the response recommended by a steadfast rule, having a credence of 1.0 in R (i.e. not to care about Bob

disagreeing) are mixed according to the confidence Alice has in R , 0.85, and steadfastness, 0.15. But then, R gives inconsistent advice: it has (directly) recommended a degree of belief of 0.85 in R , and (indirectly) recommended a degree of belief of 0.92.

This is the problem of self-undermining in a generalized form. The difference to self-undermining in a framework of absolute beliefs is that the rule does not advise to abandon itself and use the competing rule which your peers endorse, but instead advises to mix in the response of the competing rule. Nevertheless, this suffices to give rise to inconsistency. Therefore, it seems like (ED) can only motivate a rule which still faces the problem of inconsistency.

It might be replied that this objection fails because (ED), in fact, does not necessarily lead to the rule R . Rather, (ED) might be considered to lead, just as plausibly, to the rule R' : always give equal weight to your credence and that of your peer, but if your peer is *less confident* in conciliationism than you, be steadfast. Such a rule would clearly be self-exempting also in cases where Bob does not reject conciliationism, but is a bit less confident in it than Alice. Therefore, the fact that R is only partially self-exempting does not show that (ED) can only motivate a partially self-exempting rule.⁹

However, how would one derive R' from (ED)? As formulated above, epistemic deference on the reasoning level was understood as not relying on premises that the peer *rejects*. Thus, if Bob *accepts* conciliationism and is only a bit less confident in it than Alice, Alice would, by definition, not be non-deferential on the reasoning level when she used conciliationism as a premise in her reasoning. Therefore, there is no trade-off between deference on the reasoning level and deference on the credence level, so clearly she has a reason to show deference on the credence level and diminish her confidence in conciliationism. But then, R' , saying that in such cases Alice should be steadfast, cannot be justified on the basis of (ED).¹⁰

Another possible reply to my objection would be that epistemic deference on the reasoning level has to be redefined, now that a framework of relative beliefs is used. It is necessary to stop sticking to the original definition of deference on the reasoning level, which is tailored to absolute beliefs and therefore makes deference on the reasoning level come into play only if the peer *rejects* conciliationism. Maybe modifying (ED) by generalizing the definition of epistemic deference on the reasoning level to a framework of relative beliefs could lead to a new version of (ED) which can motivate a rule that is self-exempting in all cases, such

as R' ?

The straightforward generalization of deference on the reasoning level would be that it means not to give greater weight to premises than the confidence your peer has in them. This entails the original requirement of deference on the reasoning level, i.e. not to rely on premises which the peer rejects: if Bob rejects conciliationism, then the only way for Alice not to rely more on conciliationism than Bob's confidence in it is not to rely on it at all. However, in contrast to the original definition, the generalized version gives deference on the reasoning level some force even if Bob does not reject conciliationism, but only has more doubts about it than Alice. Then, if Alice adjusted her credence by equal weighting, she would reason as if conciliationism was certainly true. She would thereby give a higher weight to conciliatory lines of reasoning than Bob's degree of belief in conciliationism. Therefore, using the generalized definition of deference on the reasoning level, Alice would be non-deferential to Bob if she adjusted her belief by equal weighting. Furthermore, the two levels trade off: the more deferential she is on the reasoning level (i.e. the less she relies on premises to a higher degree than Bob's degree of belief in them), the less deferential she is on the credence level. But, one might be tempted to think, if the two levels trade off, then (ED) does not give Alice a reason to show deference on the credence level, as we have seen in the case of Bob rejecting conciliationism. And this would mean that if the notion of deference on the reasoning level is generalized as suggested, (ED) can justify self-exemption even in cases where Bob does not reject conciliationism, but only has less confidence in it than Alice.

However, while it is true that under the generalized notion of deference on the reasoning level the two levels trade off, (ED) still motivates to reduce the credence in conciliationism. The point is that a trade-off between the two levels does not automatically guarantee that (ED) does not require Alice to show some deference on the credence level. In the case of Bob rejecting conciliationism, (ED) did not motivate showing deference on the credence level because, *in that case*, showing full deference on the reasoning level meant showing no deference on the credence level—and, as none of the two levels is privileged, showing full deference on the reasoning level (i.e. no deference on the credence level) was just as good as showing full deference on the credence level. But in the case in which Bob has, for example, a credence of 0.7 in conciliationism, matters are different. Even if Alice is fully deferential on the reasoning level, she can still give conciliatory reasoning a weight of 0.7 and hence

also show some deference on the credence level. From the perspective of (ED), this is clearly preferred to a response in which she does not rely on conciliatory lines of reasoning at all and hence shows no deference on the credence level. Such a response would do worse in terms of deference on the credence level, while not doing better in terms of deference on the reasoning level. Full deference on the reasoning level, in this case, does not entail *no* deference on the credence level. Therefore, even if deference on the reasoning level is broadened so that a similar trade-off between the two levels occurs in the case in which Bob does not reject conciliationism, (ED) still requires Alice to lower her credence in such cases. Although it might not give her a reason to favour any particular of the possible responses between giving conciliatory lines of reasoning a weight of 0.7 and giving them a weight of 1.0, it does favour any such response over a response in which she shows no deference on the credence level.¹¹ In other words, (ED) does give Alice a reason to show *some* deference on the credence level, and hence lose some confidence in conciliationism. This suffices to make the rule motivated by (ED) inconsistent.

To sum up, (ED) cannot motivate a fully self-exempting conciliatory rule. Even if Pittard was right that (ED) is what motivates conciliationists, a fully self-exempting conciliatory rule would still be arbitrary from the perspective of the concerns motivating conciliationism. Even if (ED) can be given a proper justification, consistent, full self-exempting conciliationism would still be unjustified. The rule which might be justified on the basis of (ED) is a rule which recommends steadfast responses in disagreements with peers who reject conciliationism, but not with peers who have small doubts about conciliationism. Such a rule, besides seeming rather unintuitive *prima facie*, faces the same problem of self-undermining that was hoped to be overcome by self-exemption.

Notes

- 1 Of course, the consistency of arguments for a rule on disagreement can trivially be ensured by weakening their conclusions so that there are no situations for which they yield contradictory statements about what would be rational in that situation: simply restrict the scope of their conclusions to disjoint sets of situations. But then, while the conclusions are now consistent, the premises of the arguments, taken together, would still be inconsistent because the stronger, contradictory conclusions still seem to follow from them. Furthermore, weakening the premises of the different arguments by restricting them so that they apply to disjoint sets of situations is not a viable solution either. This is because it is, *prima facie*, plausible that if one accepts the restricted version of the premises,

then one should also accept the original, unrestricted premises. Put differently, although the set of arguments is not logically inconsistent any more, holding that the restricted premises are true while denying that the unrestricted premises are false itself requires an argument.

- 2 It might seem as if Alice still relies on premises which Bob rejects, even if she reasons to the conclusion that she should not adjust her credence. Would she not justify her steadfastness by the importance of deference on the reasoning level, a premise which Bob denies? I think Pittard could reply to this objection by pointing to the following possible reasoning that Alice could go through. She could use the premises that (1) the concern for deference does not give her a reason to adjust her credence, and that (2) no other (rational) concern gives her a reason to adjust her credence. From these two premises, she can conclude that she does not have a reason to adjust her belief and thus should not adjust it. Bob neither rejects (1) nor (2). Therefore, Alice can reason to the conclusion that she should be steadfast without using premises which Bob rejects.
- 3 One might wonder whether it is correct to say that deference does not give Alice a reason to reduce her belief. Is it not more plausible to say that her concern for deference on the reasoning level gives her a reason not to adjust her belief, but her concern for deference on the credence level gives her a reason to adjust her belief by equal-weighting? Then, Alice would have one reason for no credence reduction, and one reason for a large credence reduction. But if that was true, and even under the assumption that both of these reasons are equally important, then it might look like it would be rational for her to reduce her credence a bit. I suppose Pittard would respond by emphasizing that deference is conceived of as the general concern for giving weight to the views of your peers. It is not that Alice has two separate concerns—deference on the credence level and deference on the reasoning level—which independently generate reasons to prefer certain responses to others. Rather, his point is that whether the (single) concern for deference leads to a reason to prefer one response over another depends on how well the responses do both on the reasoning level and on the credence level. And if two responses only make different trade-offs between the two levels, the concern for deference does not produce a reason to prefer one over the other. Hence, the concern for deference does not generate a reason for Alice to change her belief, nor a reason not to change her belief. It simply generates no reason at all with respect to her degree of belief in conciliationism, and thus it appears perfectly rational for her to rely on whatever other reasons she has to believe in conciliationism.
- 4 Schwitzgebel [6] argues for this intuition in a context where X is the only source of evidence one has: “Suppose your only evidence about the outside world were through video feeds. If you were to discover video-feed evidence that the video feeds were unreliable, the proper response would not be to discard that evidence as self-defeating and retain high confidence in one’s feed. It would be surprising if it were generally the case that evidence that one is in a risky epistemic position somehow bites its own tail, disappearing through self-defeat to leave certainty in its place.” I think that this intuition is preserved even if video-feeds are just one out of many sources of evidence. Of course, in this case you can get evidence about the unreliability of video-feed evidence via other channels, too, and you might give these more weight than such evidence coming from the video-feed channel itself, but still it would be irrational to ignore evidence from the video-channel suggesting that it is unreliable.

- 5 As an aside, intuitions might be that even if the evidence was directly self-defeating, it would be rational to diminish one's confidence in other evidence generated by the same mechanism because the fact that the mechanism generates self-defeating evidence indicates the unreliability of the mechanism in general.
- 6 In fact, it seems that no conciliatory view can propose steadfast responses to disagreements about disagreement while endorsing the evidence-based perspective that many conciliationists invoke to justify their views. My considerations suggest that from this perspective, self-exemption seems like an irrational response to evidence. Thus, self-exempting conciliationists, irrespective of whether they endorse (ED) or not, either need to justify conciliatory responses in other cases without appealing to such evidence-based arguments, or somehow reconcile these arguments with self-exemption.
- 7 Pittard restricts his discussion to this case: "We can imagine that my initial credence for conciliationism is close to one, whereas my disputant's well-considered credence for conciliationism is close to zero" [5, p. 8].
- 8 Here, to have a degree of belief of q in a rule X means to have a degree of belief of q in the proposition that it is rational to follow X .
- 9 Thanks to an anonymous referee for pointing out this reply to my objection.
- 10 As an aside, note that if it can be made plausible, contrary to my arguments in Section 3, that the evidence obtained in disagreement about conciliationism does not require the slightest adjustment of one's credence in conciliationism, then it might be possible to give a direct argument for R' from the evidence-based perspective, not using an intermediate principle like (ED) at all.
- 11 As an aside, full deference on the reasoning level, with the maximal amount of deference on the credence level compatible with it, would result in a credence of $0.7 * 0.85 + 0.3 * 1.0 \approx 0.9$, no deference on the reasoning level in a credence of $1.0 * 0.85 = 0.85$.

Simon Bleszenohl

University of Tübingen

<simon.bleszenohl@student.uni-tuebingen.de>

References

- [1] David Christensen. "Disagreement as Evidence: The Epistemology of Controversy". In: *Philosophy Compass* 4.5 (2009), pp. 756–767.
- [2] David Christensen. "Epistemic Modesty Defended". In: *The Epistemology of Disagreement: New Essays*. Ed. by David Christensen and Jennifer Lackey. Oxford University Press, 2013, pp. 77–97.

- [3] Adam Elga. “How to disagree about how to disagree”. In: *Disagreement*. Ed. by Richard Feldman and Ted Warfield. Oxford University Press, 2010, pp. 175–186.
- [4] Thomas Kelly. “The Epistemic Significance of Disagreement”. In: *Oxford Studies in Epistemology*. Ed. by John Hawthorne and Tamar Gendler. Vol. 1. Oxford University Press, 2005, pp. 167–196.
- [5] Pittard. “Resolute Conciliationism”. In: *The Philosophical Quarterly* (forthcoming).
- [6] Eric Schwitzgebel. “1% Skepticism”. Draft.
- [7] Brian Weatherson. “Disagreements, Philosophical and Otherwise”. In: *The Epistemology of Disagreement: New Essays*. Ed. by David Christensen and Jennifer Lackey. Oxford University Press, 2013, pp. 54–76.
- [8] Ruth Weintraub. “Can Steadfast Peer Disagreement Be Rational?”. In: *The Philosophical Quarterly* 63.253 (2013), pp. 740–759.

