Resolute Conciliationism

John Pittard

Abstract: "Conciliationism" is the view that disagreement with apparently qualified disputants gives us a powerful reason for doubting our disputed views, a reason that will often be sufficient to defeat what would otherwise be strong evidential justification for our position. Conciliationism is disputed by a number qualified philosophers, a fact which has led many to conclude that conciliationism is self-undermining. Against this conclusion, I argue that the most reasonable conciliatory norm is one that can prescribe conciliation in disagreements over most matters, but not in disagreements concerning the norm itself. Adam Elga has argued for such a self-exempting view, but he fails to properly characterize the arbitrariness objection facing self-exemption and does not provide the resources for meeting it. After laying out the arbitrariness objection in its full force, I go on to answer it by showing that self-exemption is not an ad hoc attempt at self-protection, but is rather a natural consequence of taking into account all the skeptical pressures generated by a given disagreement.

Most of us can point to some proposition that we believe with confidence despite the fact that many people who seem as qualified as ourselves to assess the proposition's plausibility hold it to be false. According to "conciliationists," such confidence in the face of informed disagreement is *prima facie* irrational: if dispute-independent considerations strongly support the conclusion that my disputants are just as qualified and well-positioned to assess the disputed proposition as those on my side of the dispute, then (absent mitigating factors) I ought to adopt a level of confidence that gives significant weight to the views of my disputants. But this conciliatory position seems to impugn itself. For however strong the arguments for conciliationism may be, many apparently

¹ This characterization of "conciliationism," which is most directly inspired by Christensen (2011), papers over various complicating factors acknowledged by conciliationists (e.g., questions concerning the independence/dependence of the disputants' opinions) as well as differences between competing conciliatory proposals. These details will not be important here. For arguments in favor of conciliatory positions, see, for example, (Feldman 2006; Christensen 2007; Elga 2007; Kornblith 2010). The term "conciliationism" is taken from Elga (2010), though I use the term in a more inclusive way than he does.

qualified and informed philosophers remain unconvinced (e.g., Bergmann 2009; Kelly 2010; Lackev 2010; Kvanvig 2011). It would seem, then, that according to the conciliationists' own theory, they cannot reasonably maintain a confident belief in this theory and perhaps ought to abandon it altogether. What's worse, some have argued that a conciliatory norm that can impugn itself in this way will in some situations issue inconsistent advice and is for this reason positively incoherent (Weatherson 2007; Elga 2010). If this is right, then even a modest conciliatory view will be unable to command rational assent if it is so much as possible for that view to be applied to itself. One response to these pressing objections to conciliationism would be to endorse a conciliatory norm that requires conciliation in most disagreements but not in disagreements where conciliationism itself is at issue. Such a self-exempting norm would seem to avoid problems associated with actual or potential self-undermining. But restricting the scope of conciliationism in this way seems unacceptably arbitrary: why, after all, should conciliationism be the sole topic where disagreement has no skeptical implications?² Adam Elga (2010) has recently attempted to defend self-exempting conciliationism against this arbitrariness objection. Unfortunately for the conciliationist, Elga fails to properly characterize the arbitrariness objection, and his discussion ultimately does not provide the resources for meeting it. But I will argue that, despite Elga's failure to meet the objection, he is nonetheless correct in his claim that self-exempting conciliationism is not unacceptably arbitrary. As I will seek to show, the case for self-exemption can be fully reconciled with the principles and concerns that motivate conciliationism; indeed, the basis for self-exemption can be explained in terms of the disagreement-generated skeptical pressures posited by the conciliationist.

² Self-exemption is considered and dismissed as unacceptably arbitrary by, e.g., Thune (2010, 17).

After addressing the arbitrariness worry in my positive argument for self-exemption, I will go on to address two other objections to my position. The first objection is that my argument, if translated to other contexts, would seem to imply that no self-defeat argument presents a serious worry to any epistemic norm, since all self-defeat worries could be avoided simply by adding a self-exemption clause to the potentially self-defeating norm; yet clearly some self-defeat arguments do present a serious worry, one that cannot be overcome by a facile appeal to selfexemption. In response to this objection, I show that my argument for self-exemption in the context of conciliationism does not generalize in a problematic way. Specifically, the strategy I employ will not successfully translate to contexts where the norm charged with self-defeat is one whose justification must be accounted for in terms of the norm itself (as will be the case with any norm that characterizes a complete theory of justification). The second objection is that the "resolute conciliationism" I defend here commends a degree of steadfastness with respect to conciliationism that is objectionably immodest and intuitively absurd. In response, I argue that the proponent of a self-exempting conciliatory norm need not be committed to an immodest intransigence when that norm is disputed. This is because a reasonable conciliationist will be committed to not just one but to a whole host of conciliatory norms; and even if all of these norms are individually self-exempting, in a situation where one of the stronger and more demanding conciliatory norms is overwhelmingly disputed, the conciliationist's belief that this norm is correct will likely be undermined by a different, weaker conciliatory norm to which the conciliationist is even more firmly committed. My defense of self-exemption thus allows for some degree of modesty with respect to individual conciliatory norms while also establishing the

possibility of maintaining a coherent and robust conciliatory commitment in the face of significant disagreement.

1. ELGA ON THE "SELF-UNDERMINING PROBLEM" AND SELF-EXEMPTING CONCILIATIONISM

I begin by summarizing Elga's presentation of what he calls the "self-undermining problem" facing unrestricted conciliationism and assessing his argument for the acceptability of self-exempting conciliationism.³ While Elga's defense of self-exemption is ultimately unsuccessful, an evaluation of his argument provides an opportunity to clarify just what a successful defense of self-exemption must accomplish and sets the stage for my own argument to follow.

1.1 The self-undermining problem

Elga claims that "one's view on disagreement is part of one's *inductive method*," where an inductive method is "one's fundamental method for taking evidence into account," describing what one ought to believe in response to various possible experiences (2010, 179–80). What makes an inductive method *fundamental* is that its application "is not governed or evaluated by any other method" (185, f.n.). According to Elga, any inductive method that includes a conciliatory view that can apply to itself will, upon encountering the relevant type of disagreement over this conciliatory view, end up recommending the adoption of some other competing inductive method. But any inductive method that recommends some competing

³ Note that Elga uses 'conciliationism' to refer only to *unrestricted* conciliatory policies; I use the term to apply to a broader array of views that may include self-exempting as well as unrestricted policies.

inductive method will offer inconsistent recommendations.⁴ To see this, suppose that *II* is my inductive method and that it requires conciliation in all disagreements with apparent "epistemic peers." It is possible that someone who appears to be my epistemic peer with respect to the epistemology of disagreement might disagree with conciliationism, advocating instead some opposed "steadfast" view. Because *II* is conciliatory, it would recommend that, upon encountering this disagreement over conciliationism, I lower my credence for conciliationism and increase my credence for the opposed steadfast view. But this amounts to recommending that I adopt some *other* inductive method, *I2*, that is less conciliatory in its approach to disagreement than *II*. Clearly, *II* and *I2* will differ in some of their recommendations, since they recommend different doxastic responses to peer disagreement. There will, therefore, be some possible circumstance *C* involving peer disagreement to which *II* and *I2* offer mutually incompatible recommendations *R1* and *R2*, respectively. So if *I1* says to follow *I2*, then *I1* implicitly recommends *R2* in response to *C*. But *II* also recommends *R1* in response to *C*. Thus,

⁴ Elga's argument for this conclusion (found on pp. 179-182) is adapted from Hartry Field (2000, 131).

⁵ The term "epistemic peer," common in the disagreement literature, was first introduced by Gutting (1982). The terms is used differently by different authors, but for purposes of this paper we can say that two people are epistemic peers with respect to *p* just in case their epistemic credentials (including the quality and quantity of their evidence and their capacities for rationally assessing that evidence) make them equally qualified to assess the plausibility of *p*. And two people are *apparent* epistemic peers with respect to *p* just in case, setting aside facts about their actual views on *p*, they appear to be epistemic peers with respect to *p*.

⁶ Strictly speaking, the fact that *II* recommends that I increase my credence in a steadfast view does not entail that *II* recommends that I *follow* a steadfast policy. For it is certainly possible for *II* to recommend that I increase my credence for a steadfast view while continuing to perfectly conform to *II*. But arguably, a *reasonable* inductive method will require that my reliance on competing methods be proportioned to my credences for those methods. If this is right, then if *II* is reasonable, in recommending that I increase my credence for a steadfast view, it will also recommend that I give some weight to the prescriptions of a steadfast policy in the formation of my credences. By taking on board this additional premise (which admittedly is not uncontroversial), Elga's argument could be reframed as an argument applying to any *reasonable* inductive method: since a reasonable inductive method will require that reliance on methods be proportioned to credence for those methods, and since this requirement cannot be coherently combined with a fundamental conciliatory policy that could apply to disputes over the policy itself, we can conclude that this sort of unrestricted conciliatory policy cannot be part of any reasonable inductive method.

II is an incoherent inductive method. Elga's argument suggests that this fatal "self-undermining problem" afflicts any inductive method that includes a conciliatory requirement that can call for conciliation in disagreements over the correctness of the requirement itself.

1.2 Elga's defense of self-exempting conciliationism

A self-exempting conciliationism that did not recommend conciliation in cases where conciliationism itself is under dispute would avoid the self-undermining problem. But Elga acknowledges that self-exempting conciliatory views "seem to require arbitrary and ad hoc restrictions" (183). And if the only coherent versions of conciliationism are objectionably arbitrary, this would seem to give us a good reason for rejecting conciliationism altogether. But Elga goes on to argue that the apparent arbitrariness of self-exemption is an "illusion" and that self-exemption is in fact well motivated (184-6). Self-exempting conciliationism is not unacceptably ad hoc, according to Elga, since the requirement that conciliationism be selfexempting follows from "the completely general constraint" that "in order to be consistent, a fundamental policy, rule, or method must be dogmatic with respect to its own correctness" (185). A method is "dogmatic with respect to its own correctness" only if under no circumstances does it recommend abandoning the method. Since any fundamental epistemic method must be dogmatic in this sense if it is to be consistent, Elga thinks that it is not arbitrary for a conciliatory policy to exclude conciliationism itself from its scope. Elga goes on to suggest that the reason for self-exemption in this case is analogous to the reason that Consumer Reports would be justified in consistently and automatically rating itself the best consumer product ratings magazine: they must give themselves a superior ranking in order to be consistent with their buying

recommendations for toasters and cars and the like (which sometimes conflict with recommendations from competing magazines) (185).⁷ Similarly, a conciliatory method *must* unconditionally endorse conciliationism if it is to consistently recommend conciliation in disagreements with apparent peers on other topics.

I will develop two criticisms of Elga's defense of self-exempting conciliationism: first, his case for self-exemption relies on an unjustified exclusive focus on fundamental methods; second, he fails to properly characterize (and address) the objection that self-exemption is unacceptably arbitrary or ad hoc.

1.2.1 Criticism one: An unjustified exclusive focus on fundamental methods

Elga asserts that one's view on disagreement is "part" of one's "inductive method," which Elga has defined in such a way as to stipulate that inductive methods are *fundamental* methods. By saying that one's view on disagreement is "part" of one's inductive method, Elga means that one's inductive method says everything that one's view on disagreement says (180 f.n. 6). But I will argue that Elga is unjustified in thinking that one's view on disagreement must be fundamental in this way. And once we consider the possibility that one's view on disagreement could be provisional rather than fundamental, Elga's case for self-exemption falls apart.

In order to illustrate why Elga's appeal to the requirement that fundamental methods be dogmatic is ultimately beside the point, it will be helpful to reflect further on Elga's *Consumer Reports* example. Elga is right that if the editors of *Consumer Reports* are to consistently stand by their consumer product advice, they must not believe that the advice of *Consumer Reports* is

⁷ Elga takes his inspiration for the *Consumer Reports* example from Lewis (1971, 55).

inferior to the contrary advice of some specified competitor magazine. But from this we cannot conclude that the editors of Consumer Reports should automatically rank their publication as the most reliable consumer product magazine. As long as it is possible for the editors to become less confident in their past product recommendations, consistency does not require that they adopt a dogmatic approach to ranking consumer product magazines. To see this, suppose that *Consumer* Reports has up to this point stood by method M as the best method for ranking consumer products of any category in terms of their success in delivering specified goods and values. But now some meddlesome staff worker informs the editorial board that he has used M to rank consumer product guides according to the reliability of their consumer product rankings, and that after proper application of M he has arrived at the conclusion that the competitor magazine Consume or Die, which uses another ranking method, has rankings that are reliably better than those of Consumer Reports. Of course Elga is right that Consumer Reports cannot consistently stand by its product rankings while also ranking Consume or Die as the most reliable consumer products guide. ⁸ But this most certainly does *not* mean that the reasonable thing for *Consumer* Reports to do in response to this discouraging news is to shift their methodology from M to M*. where M* says to use M in all product evaluations except when evaluating consumer product guides (in which case M* simply says to rank first those guides which use M*). While this example is under-described, the most natural reaction to the example is to judge that the editorial board should retract its previous advice and either change their ranking methodology or close up shop. Surely it would be utterly unconvincing if the editorial board justified their adoption of M*

-

⁸ I'm assuming that *Consumer Reports* and *Consume or Die* sometimes rank the same products, and that M says that *Consume or Die* is superior even when the same products are being ranked.

on the basis of "the completely general constraint" that "in order to be consistent, a fundamental policy, rule, or method must be dogmatic with respect to its own correctness." For while adopting the dogmatic M* restores rational consistency, there are likely other rationally consistent options available (like retracting former rankings and substantively changing the ranking method) that are less dogmatic and *precisely for that reason* seem much more reasonable.

By extending the Consumer Reports example and imagining the adoption of M* in response to the self-undermining of M, the dogmatism Elga commends begins to look patently absurd. But this does not mean that we should look for some flaw in Elga's argument that inductive methods must be dogmatic in order to be coherent; I think this argument is correct. Rather, the problem is with Elga's exclusive focus on inductive methods, which for Elga are *fundamental* methods. Some reflection on the elaborated *Consumer Reports* example helps us to see that it can be perfectly reasonable to endorse some method of belief formation without thereby holding that method as a fundamental commitment. Before the meddlesome staff worker attempted to use M to rank consumer guides according to their reliability in ranking products, surely the editors could have been reasonable in holding the following beliefs: (i) M is the best method for ranking consumer products according to specified criteria; (ii) should M be used to rank consumer product guides according to their reliability in ranking products, it would deliver the verdict that guides using M are superior to other guides; (iii) nonetheless, it is epistemically possible that applying M in this way could deliver the verdict that guides *not* employing M are superior to those that do employ M; and (iv) in the event that using M to rank guides delivered this unexpected verdict, the reasonable thing to do would be to abandon the belief that M is the best

method for ranking consumer products. Clearly an editor who affirmed these four beliefs about M could be described as a proponent of M. But such an editor would *not* have M as part of her fundamental method, nor could M be derived from her fundamental method. Since her commitment to M is conditional on her not receiving evidence indicating that M has issued a self-undermining verdict, M is a *provisional* rather than fundamental commitment. And once we allow that proponents of M may hold M as a provisional rather than fundamental commitment, there is no reason to suppose that proponents of M have an incoherent inductive method on account of M's not being dogmatic (unlike M*). For in contexts where M self-undermines, those who held M provisionally are no longer committed to M, and thus they are not committed to inconsistent courses of action in such contexts. So the requirement that inductive methods be coherent does not provide a sufficient reason for giving up M and adopting M* in its place, since the coherence requirement can also be met by holding M in a non-fundamental way.

This discussion suggests that Elga is too quick to suppose that a conciliationist's (fundamental) inductive method will say everything that her favored conciliatory norm says. And if it is possible for the conciliationist to endorse a conciliatory norm in a non-fundamental way, then the requirement that inductive methods be coherent does not by itself support the conclusion that conciliationists ought to endorse a dogmatic, self-exempting conciliationism. For before learning that there actually *is* disagreement over conciliationism, there appear to be at least two ways to be a conciliationist without having an incoherent inductive norm: adopt a self-exempting conciliatory norm, or have a non-fundamental commitment to a non-self-exempting conciliatory norm. In the latter case, one's fundamental norm would not require that one fully conform to the conciliatory norm in contexts where, due to disagreement over the conciliatory norm, one has

lost confidence in the conciliatory norm. In such contexts, the fundamental norm would require only that one partially follow the conciliatory norm, weighting its prescriptions in proportion to one's credence for the proposition that the conciliatory norm is correct. Since Elga's incoherence argument relies on the assumption that a conciliationist has a fundamental commitment to a conciliatory norm and is therefore committed to fully conforming to that norm in all contexts, the argument does not apply to someone who has a non-fundamental commitment to conciliationism in the way just described.

Reflecting on the above elaboration of the *Consumer Reports* example suggests that adopting a dogmatic, self-exempting method may in some circumstances be less reasonable than provisionally committing to a method that could self-undermine and abandoning or weakening this commitment in contexts where it does self-undermine. The elaboration of the *Consumer Reports* example also poses a challenge to me, since I will have to show why my defense of self-exemption in the case of conciliationism does not imply that the adoption of M* would be a reasonable course of action. I will return to this challenge in §3.

1.2.2 Criticism two: A failure to properly characterize (and address) the arbitrariness objection

My second and more important objection to Elga's defense of self-exemption is that it does not actually address the legitimate concern expressed by the charge that self-exemption is unacceptably "arbitrary." To show why, we will need to spell out this concern more explicitly. The arbitrariness objection is *not* that the conciliationist has absolutely no reason for adding self-exemption to her theory, nor is the objection that the reason for self-exemption is not general

enough. For it is obvious that the conciliationist has a reason for adding a self-exemption clause: namely, avoiding epistemic self-defeat and (possibly) incoherence. And avoiding self-defeat and incoherence are essential concerns for *any* position. Rather, the arbitrariness objection is that *from the perspective of the concerns motivating conciliationism*, there is no reason to restrict conciliatory requirements to topics other than conciliationism. Arguments for conciliationism seek to show that disagreements generate a kind of skeptical pressure, and if self-exempting conciliationism is correct, there must be an explanation for why this skeptical pressure does not arise or is somehow overcome in disagreements over conciliationism itself. But the fact that adding a self-exemption clause would preserve the coherence of conciliationism in no way explains why disagreements over conciliationism do not generate skeptical pressure. The "high level" concern of coherence does not rebut or undermine the "ground level" reasons for disagreement-based skepticism, reasons that seem to apply to disagreements over conciliationism as much as to any other type of disagreement.

Of course if we knew that some form of conciliationism was correct and that (as claimed by Elga) a self-exempting conciliationism is the only coherent form of conciliationism, then we would have good reason for accepting self-exempting conciliationism. But even so, we would still be left with the puzzle of explaining at the "ground level" why disagreements over conciliationism do not generate skeptical pressure. And given that we do *not* know that conciliationism is correct, the lack of such a ground-level explanation for self-exemption, taken together with the putative incoherence of unrestricted conciliationism, amounts to an objection against conciliationism as a whole.

The challenge to the conciliationist, then, is to explain why the skeptical pressures typically presented by peer disagreement are absent or overcome in controversies over conciliationism.

Elga does not meet this challenge. He mistakenly understands the arbitrariness objection against self-exemption as being that self-exemption is arbitrary *absolutely*, when instead the objection is that self-exemption is arbitrary from the point of view of the ground-level concerns used to motivate conciliationism in the first place.

2. THE CASE FOR SELF-EXEMPTION

I will offer two independent and mutually supporting arguments for the claim that self-exempting conciliationism is not unacceptably arbitrary. One argument seeks to articulate the "ground-level" basis for self-exemption, explaining why the skeptical pressure that is normally generated by disagreement is lacking in disagreements over conciliationism. But however correct this explanation may be, it is likely to be persuasive only to the extent that one feels the need for such an explanation. As long as one thinks that we can reasonably get by without endorsing any suspiciously self-exempting norms, one is likely to remain suspicious of any argument for self-exemption, whether or not there is an evident flaw in the argument. After all, the intuition that self-exempting conciliationism is illegitimately arbitrary is extremely strong. Given this strong intuition, it is only natural to treat arguments for self-exemption with skepticism and to resolve to steer clear of norms with suspiciously convenient restrictions in scope.

In order to preemptively block such a reaction, I also offer a preparatory argument that shows that there are situations where self-exempting norms can be avoided only at great rational cost.

Even if one can reasonably deny conciliationism and thus avoid any pressure to endorse a self-

exempting norm in the case of peer disagreement, there are other cases that are structurally similar to peer disagreement where all of the alternatives to a self-exempting norm are decidedly irrational. In these cases, it seems that a self-exempting norm is by far the best option, however arbitrary such a norm may at first appear.

I will first present the argument that self-exemption must be endorsed in some situations that are structurally similar to that of peer disagreement, and only then proceed to give my account of why self-exemption is not arbitrary. It is hoped that once the reader sees the need to endorse self-exempting norms in some contexts, she will be less resistant to the idea that self-exemption can be given a proper ground-level explanation, and thus more receptive to the explanation I provide.

2.1 Why we all must endorse self-exempting norms

In this section, I develop an example of a skepticism-inducing situation that is structurally similar to peer disagreement but where it is *obvious* that whatever norm characterizes the correct response to the situation will be a *skeptical* norm (as opposed to a "steadfast" norm). I then show that, in the context of the example, a skeptical norm that is *not* self-exempting is irrational (even if that norm is a provisional rather than fundamental commitment). We are thus led to the conclusion that, in this situation that is structurally similar to peer disagreement, the most rational approach would be to endorse a self-exempting norm. By showing that we all must endorse a self-exempting norm in at least some situations, this argument renders apparent the need for an explanation of why self-exemption is not arbitrary. I supply such an explanation in §2.2.

2.1.1 The game of Doubt Inducement

Evil Angel enjoys playing a game that she calls "Doubt Inducement." The game of Doubt Inducement capitalizes on Evil Angel's astonishing ability to manipulate people's beliefs by manipulating the evidence to which they are exposed and by directly manipulating their neural circuitry. The game works as follows. First, Evil Angel chooses a hapless victim (usually a young and impressionable one), and then she proceeds to choose one hundred propositions that she knows are propositions that she can bring the victim to either confidently believe or confidently disbelieve through various manipulation techniques. Then, for each proposition, she flips a coin. The coin is fair and the result of the flip is not under Evil Angel's control. If the coin comes up heads, Evil Angel will bring it about that the victim comes to confidently believe the proposition. If the coin comes up tails, then Evil Angel will bring it about that the victim comes to confidently believe the *negation* of the proposition. In either case, it will seem to the victim that his belief concerning the proposition is based on very good reasons. Of course, in some cases the beliefs may in fact be based on good reasons. For example, if some true proposition p is on the list, and the coin flip for p comes up heads, then possibly no manipulation would be required if the victim would get good evidence for p without any intervention by Evil Angel; and any manipulation that is required might just involve leading the victim to encounter legitimately good evidence for p. But even beliefs that are subject to misleading manipulation will seem to the victim to be very well-justified beliefs. Moreover, the manipulated beliefs will all be stable beliefs that survive close inspection and reflection. This kind of manipulation may of course require that Evil Angel manipulate certain other beliefs (e.g. beliefs about what evidence one possesses), but Evil Angel does not engage in more manipulation than is necessary to produce

the desired effect, and she is careful not to choose propositions that might force her to massively deceive her victims or bring about significant cognitive impairments in order to bring about the desired belief. (Using the imagery of then-candidate Barack Obama, she uses a scalpel in her operations, never a hatchet.) Finally, it is never the case that the truth of a proposition on the list is in any way dependent on the result of the coin flip. So it is never the case that Evil Angel makes it the case that p (or p) because the coin flip requires her to bring about a belief that p (or that p) in her victim. Rather, the proposition's truth value is independent of the coin flip, and Evil Angel takes whatever steps are necessary to bring about the desired belief by manipulating the victim's evidential environment and his neural circuitry.

That is Phase 1 of Doubt Inducement. In Phase 2, which occurs when the victim is a reflective adult, Evil Angel makes a deal with Honest Angel, an angel who is unable to say false things. Evil Angel offers to tell Honest Angel the identity of one of Evil Angel's victims and to give Honest Angel a deck of one hundred cards bearing the one hundred propositions that Evil Angel selected for that victim. But in order to receive this information, Honest Angel must agree to the following terms. Honest Angel must explain the game of Doubt Inducement to the victim, inform the victim that he is one of Evil Angel's victims, and explain that the propositions on the cards together constitute the one hundred propositions selected for him by Evil Angel. Honest Angel must do her best to make the victim very confident that she is telling the truth, but under no circumstances can she intentionally give clues as to which of the propositions in the deck are true and which are false. Honest Angel should then proceed to shuffle the deck, draw

⁹ More accurately, *sentences* expressing these propositions will be written on cards, but we can assume that the sentences are written in such a way that the proposition each sentence expresses is unambiguous, and I will speak of cards with propositions rather than sentences.

one card at a time, and hand each card over to the victim. Honest Angel is a helpful angel, so she always agrees to Evil Angel's terms and carries them out. Given her inability to lie or learn the identity of Evil Angel's victims without agreeing to Evil Angel's terms, the best she can do to help the victims is to play along with Evil Angel's scheme.

Two further clarifications of the terms of Doubt Inducement are needed. First, Evil Angel is careful to ensure that her victims, even after reading the list of 100 propositions, will not be able to identify some pattern in how they have come to acquire and justify their beliefs regarding these propositions that would enable them to separate the false beliefs from the true ones. The way that the false beliefs were learned and supported won't look particularly unusual to the victim. (This doesn't mean there isn't anything unusual about the way these beliefs were acquired, but simply that Evil Angel is careful to ensure that the victim won't *notice* anything unusual. Evil Angel's ability to directly manipulate neural circuitry makes it fairly easy for her to cover her tracks.) Second, while Evil Angel manipulates what her victims think about the target propositions, she does *not* manipulate how her victims adjust their confidence levels in these propositions in light of having drawn them in a game of Doubt Inducement. So even if some victim is manipulated into believing p, that victim is still free to doubt p upon drawing a card bearing p in a game of Doubt Inducement. It's just that it will still seem to the victim that, had she not drawn p in a game of Doubt Inducement, believing p would be reasonable. So we could say that Evil Angel manipulates the victim's response to the *first-order* evidence bearing on a proposition, where the first-order evidence includes all evidence except the fact that the proposition was drawn in a game of Doubt Inducement, without manipulating the victim's

response to the second-order evidence learned in a game of Doubt Inducement. And we can suppose that this fact is made clear in Honest Angel's account of the game's terms.

2.1.2 The skeptical force of Doubt Inducement

Now, suppose that I am visited by Honest Angel and am informed that I am a victim/participant in one of Evil Angel's "games" of Doubt Inducement. As I read the propositions on the cards handed to me by Honest Angel, how should I adjust my confidence level in my beliefs regarding the truth or falsity of those propositions? One very natural view is that to the extent that I am confident that I am indeed playing a game of Doubt Inducement, I should typically assign a credence of around 0.5 to the propositions read in a game of Doubt Inducement. For if I am playing Doubt Inducement, then the levels of first-order evidential support I have for the various propositions on the cards, or at least my assessments of the evidential support for the propositions, are fixed according to a random process and thus do not correlate with the propositions' truth values. So I cannot justifiably make any appeal to the first-order evidence, since my first-order evaluations are collectively highly unreliable and there is no way for me to distinguish which beliefs are the product of deception and which are not. And since the secondorder evidence acquired in playing Doubt Inducement indicates that there is a 0.5 chance that my first-order belief is correct and a 0.5 chance that it is false, my updated credence for the proposition should typically be around 0.5. Now suppose I am uncertain as to whether I am in fact playing Doubt Inducement, or whether this whole thing is a farce and in fact my beliefs have not been manipulated. In this case, it seems that, to the extent that I think it likely that I am playing Doubt Inducement, I should typically favor a credence of around 0.5 for drawn

propositions, and to the extent I that I think it likely that I am *not* playing Doubt Inducement, I should typically favor my first-order credences. Let's call this view **GO AGNOSTIC**.

Even those who are inclined to reject conciliationism will almost certainly be inclined to accept that GO AGNOSTIC (or some similar "go agnostic" policy) is correct. GO AGNOSTIC is, after all, a very modest claim. It leaves open the possibility that I should not think that I am playing Doubt Inducement, or that when I draw propositions of a certain type, I should reduce my confidence that I am playing Doubt Inducement. It even allows that in some exceptional cases I should retain my belief concerning a drawn proposition even if I am confident that I am playing Doubt Inducement. The motivating idea behind GO AGNOSTIC is merely that if I really do think that my beliefs about a certain list of propositions have been manipulated by an extremely powerful, crafty, and deceptive angel in the way Honest Angel describes, then to the degree I think this I should *in general* move towards an agnostic stance with respect to a proposition drawn in the game. ¹⁰ Even if there is something objectionable in the way that GO AGNOSTIC

 $^{^{10}}$ [Name removed] objected to GO AGNOSTIC on the grounds that, upon drawing 1+1=2 in a game of Doubt Inducement, GO AGNOSTIC would have me respond by doubting that 1+1=2, a prescription that is clearly irrational. I have two responses. First, propositions that are as deeply embedded in my noetic structure as 1+1=2 cannot be drawn in a game of Doubt Inducement. It was stipulated that Evil Angel chooses propositions that will not force her to rely on massive deception or significant cognitive impairment in order to bring about the desired belief. Since massive deception and significant impairment would be required to deceive me about propositions that are fundamental to my doxastic system, 1+1=2 will not be drawn in a game of Doubt Inducement. GO AGNOSTIC therefore does not say what to do in a similar game where such propositions might be drawn. Second, even if the terms of Doubt Inducement did not exclude the possibility of my drawing 1+1=2, my defense of a self-exempting version of GO AGNOSTIC can explain why I should not doubt 1+1=2 upon drawing it in a game like Doubt Inducement. As I will argue below, a fully articulated version of GO AGNOSTIC ought to be self-exempting in that it ought to exclude from the scope of its "agnosticism" prescription any drawn proposition p that is such that my "going agnostic" with respect p would rationally require that I reduce my confidence in GO AGNOSTIC. This means, of course, that drawing GO AGNOSTIC should not lead me to doubt GO AGNOSTIC. But 1+1=2 is also a proposition that I cannot reasonably doubt without also doubting GO AGNOSTIC. For if I seriously think I could be wrong about something as obvious and fundamental as 1+1=2, I have no business trusting my reasoning about the correct approach to Doubt Inducement! So it turns out that the self-exemption requirement also explains why propositions as fundamental to my doxastic system as 1+1=2 will be exempt from the agnosticism prescription that applies to other propositions.

characterizes this agnostic stance (as an "agnostic credence" of around 0.5), I take it to be fairly clear that some sort of "go agnostic" policy is correct.¹¹

2.1.3 Ruling out unrestricted Doubt Inducement norms: a preliminary case

While GO AGNOSTIC seems correct, or at least *approximately* correct, GO AGNOSTIC makes a claim only about what I should *typically* do in a game of Doubt Inducement and therefore does not specify whether there are exceptional cases where one should *not* "go agnostic" with respect to the drawn proposition (despite confidence that one is playing Doubt Inducement). And while it is intuitively clear that going agnostic is *in general* the reasonable response if one is confident that one is playing Doubt Inducement, it is not clear that this is *always* the reasonable response.

And I shall now argue that there are in fact exceptions. In particular, I will argue that I should not go agnostic with respect to a drawn proposition if my doing so would rationally require that I reduce my confidence in GO AGNOSTIC. This means that GO AGNOSTIC itself, along with propositions that implicate GO AGNOSTIC in the relevant way, will be exceptions to the general prescription characterized by GO AGNOSTIC. It turns out that a fully spelled-out epistemic norm

-

Against GO AGNOSTIC, some might maintain that the proper response to drawing *p* in a game of Doubt Inducement is to adopt a more thoroughly agnostic stance by refusing to assign any credence whatsoever to *p*. (Thanks to Dustin Locke for raising this possibility.) I doubt that this is a superior policy. Suppose that the parameters of Doubt Inducement were changed, so that there is a 0.8 chance that the victim's manipulated belief concerning a drawn proposition is true, and only a 0.2 chance that it is false. It seems that if the "withhold credence" strategy is the best approach to the original version of Doubt Inducement, then it will also be the best approach to this revised version. And if withholding is the appropriate response in both versions of the game, then the level and type of doubt that is rationally required does not differ across the two versions of the game. Yet surely a less extreme level of doubt is called for in the revised version than in the original version, since the probability of false belief is greater in the latter. But even if this objection is not conclusive, my argument does not ultimately depend on whether this "withhold" policy or GO AGNOSTIC is the superior approach. As I argue in footnote 13, self-exemption will be required in either case.

describing the rational responses to drawing propositions in a game of Doubt Inducement (a "Doubt Inducement norm" for short) will be a *self-exempting* norm.

Suppose that I draw a card bearing a proposition implicating my policy of GO AGNOSTIC; specifically, suppose I draw a card bearing the proposition STEADFAST, where STEADFAST is the proposition that even after drawing a proposition in a known game of Doubt Inducement, I should continue to base my credence for the proposition on the first-order evidence alone, maintaining confident belief (or disbelief) if the first-order evidence still seems to me to support my position. Should I go agnostic with respect to STEADFAST when I draw STEADFAST in an apparent game of Doubt Inducement? I will argue that going agnostic with respect to STEADFAST is irrational.

Let us consider two different Doubt Inducement norms that both affirm everything that GO AGNOSTIC affirms but that take different stances on the question of whether STEADFAST is an exceptional case. First, let SELF-EXEMPTING AGNOSTICISM be a Doubt Inducement norm that is just like GO AGNOSTIC except that it stipulates that STEADFAST and other propositions that implicate my "go agnostic" approach to Doubt Inducement are exempt from the agnosticism prescription that applies to most other drawn propositions. So according to SELF-EXEMPTING AGNOSTICISM, a game of Doubt Inducement cannot lead me to rationally doubt whether a game of Doubt Inducement should in general induce doubt. Second, let UNRESTRICTED AGNOSTICISM be a Doubt Inducement norm just like GO AGNOSTIC except that it stipulates that propositions like STEADFAST do *not* present exceptions to the general prescription of GO AGNOSTIC. So according to UNRESTRICTED AGNOSTICISM, a game of Doubt Inducement *can* lead me to rationally doubt

whether Doubt Inducement poses a legitimate skeptical worry (and thus whether UNRESTRICTED AGNOSTICISM and similar "go agnostic" norms are correct).

I will now argue that UNRESTRICTED AGNOSTICISM is rationally incoherent, and that if I am to endorse a skeptical approach to Doubt Inducement (as I obviously should), SELF-EXEMPTING AGNOSTICISM (or some similar self-exempting norm) is the only coherent option. Importantly, unlike Elga's argument against unrestricted conciliationism, my argument in no way relies on the assumption that UNRESTRICTED AGNOSTICISM is a fundamental method that I am committed to perfectly following in all circumstances. So I will assume that, if UNRESTRICTED AGNOSTICISM undermines itself upon my drawing STEADFAST, my responses to *other* propositions drawn in the game of Doubt Inducement can be adjusted so as to be rendered consistent with my newly-lowered credence for UNRESTRICTED AGNOSTICISM. The problem that I will identify immediately results from the self-undermining of UNRESTRICTED AGNOSTICISM. So even if a game of Doubt Inducement involved only one proposition rather than one hundred, UNRESTRICTED AGNOSTICISM would still lead to incoherence in the case where the one proposition drawn is STEADFAST.

There are, it turns out, two incompatible ways of interpreting UNRESTRICTED AGNOSTICISM as it bears on the STEADFAST case: a straightforward interpretation and a more sophisticated interpretation. I will take up the more sophisticated interpretation in the next section, focusing for now only on the straightforward interpretation. According to this interpretation, upon drawing STEADFAST, I ought to go agnostic with respect to STEADFAST by adopting a credence for STEADFAST of around 0.5. The following line of reasoning suggests that this would be an irrational response. I should be agnostic with respect to STEADFAST (upon drawing it in the game) only if I am confident that GO AGNOSTIC, or some similar norm, is correct. For I have no other

grounds for increasing my credence for STEADFAST to the agnosticism range. But it is also the case that I should be agnostic with respect to STEADFAST only if I am also agnostic (or even more doubtful) with respect to the proposition that GO AGNOSTIC or some similar norm is correct. For since I know that GO AGNOSTIC and similar "go agnostic" norms entail that STEADFAST is false, my confidence in the claim that one of these norms is correct cannot rationally exceed my confidence that STEADFAST is false. And if my credence for STEADFAST is close to 0.5, then my credence for the claim that STEADFAST is false must also be close to 0.5. Thus, given an agnostic credence for STEADFAST, my credence for the claim that GO AGNOSTIC or a similar norm is correct can at most be approximately 0.5. But the implication of the above claims is that I should be agnostic with respect to STEADFAST only if (i) I am confident that GO AGNOSTIC or some similar norm is correct and (ii) I am agnostic (or more doubtful) with respect to the proposition that GO AGNOSTIC or some similar norm is correct. But the consequent of this conditional is absurd: it is irrational to both be confident in some proposition and be agnostic with respect to that very same proposition.¹² We can avoid an absurd conclusion only by denying that I should be agnostic with respect to STEADFAST. Thus, in order to avoid irrational implications, a fullyspecified Doubt Inducement norm that recommends a "go agnostic" approach to Doubt Inducement must exempt STEADFAST from the prescription of agnosticism that applies to most other drawn propositions. Let's call this argument the **no justification objection**, since the basic

¹² Note that I am using 'agnostic' in this argument as shorthand for having a credence level around 0.5. I'm *not* using 'agnostic' to mean simply lacking both the belief that p and the belief that $\neg p$. That kind of agnosticism might be compatible with being fairly confident in the truth of the proposition towards which one is agnostic. But agnosticism in the sense meant in the argument is not compatible with having a high level of confidence.

idea is that as soon as I go agnostic with respect to STEADFAST and UNRESTRICTED AGNOSTICISM, I lose any justification I may have had for this response.¹³

The general lesson of the no justification objection is that a defeater cannot defeat the very beliefs that must be presupposed for its operation as a defeater. ¹⁴ If p were to be one of my reasons for taking E to be evidence that defeats my belief that p, I would have to both believe p and not believe p: for p can rationally be a reason for me only if I believe p, and I can rationally take E to defeat the belief that p only if I do not believe p.

¹³ The policy of withholding credence assignments from propositions drawn in Doubt Inducement, which was discussed in footnote 11, will also succumb to the no justification objection if it fails to exempt STEADFAST from the general prescription of agnosticism. Suppose I endorse a withhold policy and apply this policy when I draw STEADFAST, withholding all judgments about STEADFAST's plausibility and becoming indifferent towards all possible credence values from 0 to 1. Rational consistency will require that I similarly withhold all plausibility judgments concerning the proposition that my withhold policy is correct—a proposition that I will label 'WITHHOLD.' (Why? Given the incompatibility of STEADFAST and WITHHOLD, my assigning some credence c to WITHHOLD would result in a rational upper limit of 1-c for the possible credence values for STEADFAST, and such an upper limit is incompatible with complete withholding.) But if I withhold all judgments concerning the plausibility of WITHHOLD, then I cannot rationally affirm that my withholding judgment concerning the plausibility of STEADFAST is a reasonable response to drawing STEADFAST in a game of Doubt Inducement. For the only possible basis I have for thinking that this is a reasonable response is whatever general considerations favor a withhold strategy; and I can have no opinion about the force of these considerations once I withhold judgment concerning WITHHOLD. Thus, I have no justifying reason for my newly agnostic stance. And unlike cases where I withhold plausibility judgments because I have never considered a proposition, or because I have no opinions on how to evaluate the proposition, this is surely a case where the stance of withholding stands in need of justification. For even in this agnostic state, I affirm that, setting aside the evidence of my having drawn STEADFAST, there are considerations that seem to weigh heavily in favor of WITHHOLD and against STEADFAST.

One might worry that my reasoning here would render *reductio* arguments unacceptable, since there is a sense in which they show some conclusion to be false by presupposing that conclusion. But as the proponent of a *reductio* argument, my basis for believing $\neg p$ does not include the belief that p; rather, the basis for my believing $\neg p$ is that p leads to inconsistency. Adopting p as a premise for sake of argument is just a way of demonstrating the inconsistency, and does not mean that I somehow believe $\neg p$ on the basis of p. But in the present example, robust confidence in GO AGNOSTIC would have to be part of the basis for my agnosticism towards GO AGNOSTIC, and that *is* irrational!

2.1.4 The "equilibrium defense" of UNRESTRICTED AGNOSTICISM

Once one has seen how my adopting a credence of around 0.5 for STEADFAST would undercut my justification for thinking that this is the correct response, a natural response is to consider whether there is some more sophisticated understanding of UNRESTRICTED AGNOSTICISM that would avoid these problems. And indeed, a proposal developed by Matt Weiner (2007) in a discussion of conciliationism and self-defeat points to one approach to Doubt Inducement that could plausibly be said to be in keeping with the requirements of UNRESTRICTED AGNOSTICISM while seeming to avoid the problem identified above. ¹⁵ I will now characterize this more sophisticated approach and then argue that it suffers from multiple rational defects, including a failure to escape the no justification objection.

Suppose that before drawing STEADFAST in a game of Doubt Inducement, my credence for UNRESTRICTED AGNOSTICISM is 0.98. Upon drawing STEADFAST, UNRESTRICTED AGNOSTICISM would have me move my credence for STEADFAST up to around 0.5, and rational consistency would demand that I also drop my credence for UNRESTRICTED AGNOSTICISM to around 0.5. Let's suppose that I go ahead and update my credences for STEADFAST and UNRESTRICTED AGNOSTICISM in this way. Given these new credences, I am no longer very confident that "going agnostic" with respect to STEADFAST (and UNRESTRICTED AGNOSTICISM) is the best response to drawing STEADFAST. Since my new credence for STEADFAST is on par with my new credence for UNRESTRICTED AGNOSTICISM, I should now think that a steadfast response that involves sticking by my high initial credence for UNRESTRICTED AGNOSTICISM is just as likely to be correct as my

¹⁵ Weiner's proposal is also briefly discussed by Christensen in his (2013).

current "go agnostic" response. And given this new perspective, it seems that I should change my current response by adopting credences that split the difference between my original go agnostic approach and a more steadfast approach. But of course once I have done this (adopting a credence of 0.25 for STEADFAST and a credence of 0.75 for UNRESTRICTED AGNOSTICISM), this will again occasion a shift in my evaluation of how I ought to respond to drawing STEADFAST, leading to another shift in credences, which will in turn lead to another shift in my evaluation of how I ought to respond to drawing STEADFAST, and so on and so on. While this process of shifting credences may continue indefinitely, my credence for UNRESTRICTED AGNOSTICISM will converge on an equilibrium point between 0.5 and 1 where the downward pressure generated by my credence for UNRESTRICTED AGNOSTICISM is perfectly counterbalanced by the upward pressure generated by my credence for STEADFAST. We can calculate this equilibrium credence precisely. Let *c* be my credence for UNRESTRICTED AGNOSTICISM; *c* will be stable when the following holds:¹⁶

Equilibrium equation:
$$c = c \cdot 0.5 + (1 - c) \cdot 0.98$$

Solving for c, we get $c \approx 0.66$. With a credence of 0.66 for UNRESTRICTED AGNOSTICISM and a credence of 0.34 for STEADFAST, the skeptical implications of my credence for UNRESTRICTED AGNOSTICISM is counterbalanced by the non-skeptical implications of my credence for STEADFAST, suggesting that this position is rationally stable.

¹⁶ My use of this equation to identify the convergence point is adapted from Weiner (2007).

The view just described aims to defend UNRESTRICTED AGNOSTICISM by showing that, when paired with the correct method of applying one's Doubt Inducement norm, UNRESTRICTED AGNOSTICISM leads to a rational outcome in the STEADFAST case. Let's call this method, according to which I ought to adopt (or converge on) the credences satisfying the equilibrium equation, the **equilibrium method**.¹⁷ And let's call the defense that appeals to this equilibrium method the **equilibrium defense**. The equilibrium defense does seem to enhance the plausibility of UNRESTRICTED AGNOSTICISM. For as shown in the last section, applying UNRESTRICTED AGNOSTICISM in the straightforward manner leads to an incoherent situation where I both doubt UNRESTRICTED AGNOSTICISM and fully rely on it as the basis for my "agnostic" credence of approximately 0.5 for STEADFAST. But in applying UNRESTRICTED AGNOSTICISM using the equilibrium method, the degree to which I rely on UNRESTRICTED AGNOSTICISM and STEADFAST is perfectly proportioned to my credence for those norms.

Despite this improvement, the equilibrium defense fails to vindicate the rationality of UNRESTRICTED AGNOSTICISM. Even when coupled with the equilibrium method, UNRESTRICTED AGNOSTICISM suffers from severe rational flaws. First, UNRESTRICTED AGNOSTICISM still fails to evade the no justification objection. Second, UNRESTRICTED AGNOSTICISM violates a very basic rational constraint governing the formation of conditional credences. I develop these charges in order.

¹⁷ It seems plausible to suppose that I ought to directly adopt these equilibrium credences rather than converging on them through an indefinitely long process of credence oscillations. Even if a hypothetical series of shifting credences is what *justifies* my position, perhaps I can skip over this process to the end result rather than myopically passing through each stage as though I didn't know where it was headed.

The first and most basic problem with the equilibrium defense is that, despite appearances, it succumbs to the no justification objection developed above. One might think that the problem identified by the no justification objection can ultimately be diagnosed as a mismatch between my credence for a given norm and the weight I give to that norm's prescriptions. Since the equilibrium method requires that the weight I give to a norm's prescriptions be proportional to my credence for that norm, then it would seem that my credences for both UNRESTRICTED AGNOSTICISM and STEADFAST can be rationalized in terms of (and presumably justified by) my current levels of confidence in the competing epistemic norms, so that the no justification objection is avoided.

But it would be a mistake to think that the no justification objection is avoided merely by making sure that the degree to which I rely on competing norms is proportional to my credences for those norms. To see this, consider a case where, instead of drawing STEADFAST, I draw **BELIEVE ALL**, where BELIEVE ALL is the proposition that in a game of Doubt Inducement, the rational response is to adopt a credence of 1 for all drawn propositions. In this case, UNRESTRICTED AGNOSTICISM coupled with the equilibrium method would lead me to adopt a credence of 0 for UNRESTRICTED AGNOSTICISM and 1 for BELIEVE ALL. For as my credence for BELIEVE ALL goes up and my credence for UNRESTRICTED AGNOSTICISM goes down, I become more and more confident that I ought to believe every proposition drawn in a game of Doubt Inducement, including BELIEVE ALL itself. So if *c* is my credence for UNRESTRICTED AGNOSTICISM after drawing BELIEVE ALL, the equilibrium equation in this case is as follows:

$$c = c \cdot 0.5 + (1 - c) \cdot 0$$

Solving, we get c = 0. So unrestricted agnosticism and the equilibrium method together imply that, upon drawing Believe All in a game of Doubt Inducement, I ought to believe Believe All with total confidence and disbelieve unrestricted agnosticism with total confidence.

This example illustrates in dramatic fashion how UNRESTRICTED AGNOSTICISM can require me to adopt a set of credences for Doubt Inducement norms that I cannot rationally justify even though my reliance on the competing norms are proportional to my credences for those norms. Let's suppose that I follow UNRESTRICTED AGNOSTICISM and the equilibrium method where they lead and adopt the prescribed credences of 1 for BELIEVE ALL and 0 for UNRESTRICTED AGNOSTICISM. Given my complete confidence in the correctness of BELIEVE ALL, my current credences obviously rationalize my fully complying with what BELIEVE ALL says to do upon drawing BELIEVE ALL. But just as clearly, the mere fact that BELIEVE ALL sanctions my confidence in BELIEVE ALL is not enough to justify my position. After all, antecedent to the evidence received during the game of Doubt Inducement, BELIEVE ALL seems extremely implausible to me, whether or not it is able to validate itself according to its own absurd standards. So if I am to justifiably affirm BELIEVE ALL despite its apparent implausibility, I need a justifying reason for this belief that does not itself presuppose the truth of BELIEVE ALL. Circular justification simply won't do in this case. This means that BELIEVE ALL cannot be my basis for thinking that my response to drawing BELIEVE ALL is reasonable. But the only noncircular justification that could be given for my response is that my current position is justified because that is what resulted from my following the prescriptive implications of my past

commitment to UNRESTRICTED AGNOSTICISM. And surely I can justify my current position in terms of my past commitments only if I can endorse those past commitments as reasonable. So given my credence of 0 for UNRESTRICTED AGNOSTICISM, I cannot appeal to UNRESTRICTED AGNOSTICISM in order to justify my confidence in BELIEVE ALL. Since the only non-circular basis for my position consists of past commitments that I can no longer endorse, I am unable to rationally justify the position that results from following UNRESTRICTED AGNOSTICISM.

While the lack of adequate justification is less glaringly obvious in the STEADFAST case, the same objection applies. Once I have adopted the equilibrium credences, I can rationalize my credences for UNRESTRICTED AGNOSTICISM and STEADFAST in terms of what those norms prescribe and how confident I am in these prescriptions. But of course this rationalization simply presupposes my current credences and does not provide a non-circular justification for them. And clearly, I stand in need of a non-circular reason for my significantly-adjusted credence levels for UNRESTRICTED AGNOSTICISM and STEADFAST. Is such non-circular justification available? It seems not. For the equilibrium method prescribes my current credences for UNRESTRICTED AGNOSTICISM and STEADFAST of 0.66 and 0.34 (respectively) only if I start off with a credence of nearly 1 for UNRESTRICTED AGNOSTICISM. And given my credence of 0.66 for UNRESTRICTED AGNOSTICISM, I cannot be wholly confident that a high starting credence for UNRESTRICTED AGNOSTICISM is the most rational starting point. Since I have a 0.34 credence for STEADFAST, I must have a credence of 0.34 for the claim that my high initial credence for UNRESTRICTED AGNOSTICISM was based on some sort of confusion or deficiency in evidence. True, I haven't actually *identified* this confusion or evidential deficiency. But drawing STEADFAST nonetheless leads me to think it reasonably likely that my high initial credence for

UNRESTRICTED AGNOSTICISM did result from some such deficiency. And if I think it reasonably likely that my initial commitments were rationally deficient, then I cannot justify my current position by showing that it follows from my initial commitments.

A second objection to the equilibrium defense is that whether or not we use the equilibrium method in applying UNRESTRICTED AGNOSTICISM, UNRESTRICTED AGNOSTICISM leads to violations of a very basic rational constraint governing the formation of conditional credences. We may express this rational constraint as follows:

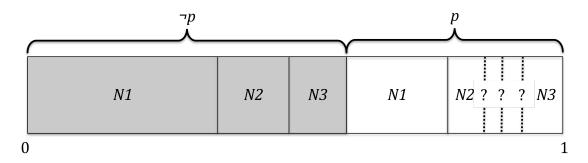
LEGITIMATE GROUNDS: If my conditional credence for h given e is defined, and if I know that p entails $\neg e$, then I cannot rationally rely on p (or "probably p") as one of my reasons for the value I assign to my conditional credence for h given e.¹⁸

LEGITIMATE GROUNDS expresses a constraint that should be rather obvious to anyone who understands what a conditional credence is. My conditional credence for h given e represents my assessment of the (subjective) probability of h on the supposition that e is true. Clearly, I cannot rely on $\neg e$ or anything that entails $\neg e$ when assessing this probability. For example, in explaining why I am confident that I will win the poker hand even if my opponent has a pair of aces, it would be irrational for me to appeal to the fact that my opponent probably does not have any aces.

Anyone who endorses a norm that in some epistemically possible situation can require me to increase my credence in a specific competing norm, as UNRESTRICTED AGNOSTICISM can, will

¹⁸ My formulation of LEGITIMATE GROUNDS benefitted from the comments of [name removed].

violate LEGITIMATE GROUNDS. Or at least they will violate LEGITIMATE GROUNDS if there are two or more epistemically possible alternatives to the self-undermining norm. To see this, suppose that norms NI - N3 are three mutually exclusive norms that collectively exhaust all possibilities, and that the following chart shows how I divide my credence between each of these norms in combination with $\neg p$ (the gray area of the bar) and with p (the white area of the bar), where p is some arbitrary proposition:



The dotted lines on the right hand side represent different possibilities for how I might divide my credence between N2 and N3 on the p side of the bar. Note that where this dividing line is drawn will determine my conditional credences for N2 and N3 given p and $\neg N1$. Thus, LEGITIMATE GROUNDS implies that N1 cannot rationally serve as one of my reasons for my drawing this line where I do. In reasoning about where to draw this line, I must suppose that N1 is false. But this means that N1 cannot give me a rational reason for thinking that, upon learning p, I ought to adopt some specific (non-zero) credence value for N2. To see this, suppose that one implication of N1 is that upon learning p, I ought to adopt credences c1, c2, and c3 for N1, N2, and N3 (respectively), where c2 is positive. If I were to endorse this prescription, I would have to draw in the missing line in a specific location, thereby fixing the values of my conditional credences for N2 and N3 given p and $\neg N1$. But as already argued, N1 cannot be a rational reason for my

drawing this line where I do. Thus, my confidence in NI (or in any propositions known to entail NI) cannot be my reason for thinking that, upon learning p, I ought to do what NI says I should do.

The upshot of this is that even if UNRESTRICTED AGNOSTICISM could supply my reason for lowering my confidence in UNRESTRICTED AGNOSTICISM upon drawing STEADFAST, it could *not* supply a rational reason for thinking that in this situation I ought to increase my confidence in STEADFAST rather than increasing my confidence in some third possible norm (like, say, RESTRICTED AGNOSTICISM). For if it were my reason for thinking this, then UNRESTRICTED AGNOSTICISM would also be my reason for thinking that, conditional on STEADFAST being drawn and UNRESTRICTED AGNOSTICISM being *false*, STEADFAST is much more likely than all of the other alternatives. And this would incoherently violate LEGITIMATE GROUNDS. Let's call this argument the **conditional credence objection**.

Summing up this section, it seems that the equilibrium defense fails to vindicate the rationality of UNRESTRICTED AGNOSTICISM. Even if we suppose that UNRESTRICTED AGNOSTICISM should be applied in the way described by the equilibrium method, UNRESTRICTED AGNOSTICISM still succumbs to the no justification objection and the conditional credence objection. It appears, then, that I cannot coherently endorse a Doubt Inducement norm that requires me to "go agnostic" with respect to propositions like STEADFAST.

2.1.5 Warming up to self-exemption

I have argued that a Doubt Inducement norm that requires me to "go agnostic" upon drawing STEADFAST leads to results that are decidedly irrational. If this is right, then if there is a Doubt

Inducement norm that is rationally unproblematic, it will either be one that typically prescribes agnosticism but that is self-exempting (along the lines of SELF-EXEMPTING AGNOSTICISM), or it will be some more "steadfast" norm that does not typically prescribe agnosticism. While a few brave souls may wish to endorse this latter option, perhaps as part of a more general position that downplays the rational significance of "higher-order" evidence, I suspect that most will find this option to be extremely implausible. Whatever one thinks about the rational implications of disagreement, it seems clear that a steadfast approach to Doubt Inducement is unreasonably sanguine. So we have good reason for thinking that if there is a rationally unproblematic Doubt Inducement norm, it will be some sort of "go agnostic" norm that is self-exempting. If such a self-exempting norm is rationally unproblematic, then we have good reason for thinking that there must be some non-arbitrary basis for the self-exemption, even if we do not yet see what that basis could be. And if there is a non-arbitrary basis for self-exemption in the Doubt Inducement case, then we should at least be open to the possibility that there is a non-arbitrary

¹⁹ The following discussion benefitted significantly from comments from [names removed].

²⁰ Brian Weatherson *might* be one "brave soul" who would commend a more steadfast approach to Doubt Inducement. In his (unpublished) (2010), Weatherson argues that my believing p on the basis of evidence E need not be defeated by (misleading) higher-order evidence that E does not support p. Weatherson suggests that maintaining my belief in the face of discouraging higher-order evidence can be rational if E does in fact support p and if I have inferred p from E by means of a basic inferential method (10-11). My inference from the fact that a thermometer says it is 60 degrees to the conclusion that it is 60 degrees does not qualify as basic, since the inference is not justified unless I have some reason to think that the thermometer is reliable. But other inferential methods, for example, the application of modus ponens, are basic in that they can confer justification without higher-order beliefs about their reliability. Given Weatherson's view that basic inferential methods may continue to confer justification in the face of discouraging higher-order evidence, perhaps Weatherson would suggest that drawing p in a game of Doubt Inducement gives me a reason to doubt my belief that p only if that belief is based on non-basic inferential methods. This strikes me as highly implausible. Suppose that in a game of Doubt Inducement I draw 100 complex claims of formal logic, all of which I believe to be true or false on the basis of long chains of logical reasoning using basic inferential rules. While the inferential methods grounding my beliefs are basic in this case, the fact that I sometimes err in my application of these methods should at least somewhat temper my confidence even when I am not being manipulated by Evil Angel. Surely, my confidence should take an even greater hit when I know that there is a very significant probability that my reasoning has been manipulated by Evil Angel.

basis for self-exemption in the structurally similar case of peer disagreement. So the above arguments ought to go a long way towards defusing any intuition-driven resistance to self-exempting conciliationism.

Matters are not so simple, however, if we are willing to entertain the possibility that there is no rationally unproblematic Doubt Inducement norm. 21 If it is possible that every approach to Doubt Inducement suffers from some sort of rational defect, then from the fact that all the alternatives to self-exemption are rationally problematic, we cannot conclude that some selfexempting approach is non-problematic (and thus that there is a non-arbitrary basis for selfexemption). But if we take this possibility seriously, the arbitrariness objection against selfexempting conciliationism will lose much of its force. The claim that self-exempting conciliationism is arbitrary and thereby rationally problematic gives us a strong reason for rejecting self-exempting conciliationism only to the extent that we are confident that there is some approach to disagreement that is *not* rationally problematic. And if there is no rationally unproblematic approach to *Doubt Inducement*, this gives us reason to doubt that there is a rationally unproblematic approach to the structurally similar case of peer disagreement. And on the supposition that there is no rationally unproblematic approach to disagreement, defending self-exempting conciliationism would not require that one show that self-exemption is not arbitrary, but would require only that one show that the arbitrariness of a self-exempting approach is less problematic than the defects that characterize other approaches to disagreement.

²¹ This possibility was raised in conversation by Keith DeRose and an anonymous referee. In his discussion of self-undermining concerns facing conciliationism, David Christensen (2013) argues that there are contexts where two or more epistemic ideals inevitably conflict, including the context where conciliationism is disputed by epistemic peers.

So if there is no rationally unproblematic approach to Doubt Inducement, then we should not be confident that there is a rationally unproblematic approach to disagreement, in which case we should not take the (putative) arbitrariness of self-exempting conciliationism to be a decisive count against it. If, on the other hand, there *is* a rationally unproblematic approach to Doubt Inducement, then we should be confident that a self-exempting approach to Doubt Inducement is not unacceptably arbitrary, in which case we should think it plausible that there is a non-arbitrary basis for self-exemption in the analogous case of disagreement. I will now attempt to describe what that basis might be.

2.2 Why self-exemption is not arbitrary

Given that apparent peer disagreement normally generates skeptical pressure that undermines one's first-order justification, why would this not be the case in disagreements over conciliationism? Why would the skeptical pressure mysteriously disappear in this case? I will now attempt to successfully meet the arbitrariness objection by answering this question.

Suppose we are convinced of the following conciliationist thesis: disagreement with apparent peers constitutes higher-order evidence that we are unreliable with respect to the disputed proposition, and, *absent any mitigating factors*, this evidence generates substantial skeptical pressure that rationally requires us to move our credence for the disputed proposition in a conciliatory direction. Having accepted this conciliationist thesis, we now turn to the question of whether a fully-articulated conciliatory policy should admit of any exceptions, cases where due to some sort of "mitigating factor," apparent peer disagreement does not generate the normal skeptical pressure. My contention is that there is such a mitigating factor in a disagreement over

conciliationism and that the skeptical pressure typically presented by disagreement does not arise in this case. In a normal case of disagreement over some belief of mine that p, the skeptical pressure generated by the disagreement pushes in a determinate direction: it raises a worry about the rationality of my belief that p, and thus creates pressure to lower my confidence that p. But in the case of disagreement with an apparent peer over my belief in conciliationism, not one but two skeptical pressures are generated, and these push in opposite directions. On the one side, the disagreement challenges the rationality of my high initial credence for conciliationism, and this generates skeptical pressure to reduce my credence for conciliationism. But on the other side, to the degree that the disagreement challenges the rationality of my belief in conciliationism, it also challenges the rationality of my taking such a disagreement to be a good reason for reducing my credence for conciliationism, thereby generating pressure on me to stick by my initial credence for conciliationism. So if in response to a disagreement over conciliationism I adopt a credence for conciliationism that is midway between my own initial credence and that of my apparent peer, I will have been quite conciliatory at the level of my final credences, but I will also have been maximally non-conciliatory at the level of my reasoning on which those credences are based. For I could justify the adoption of this intermediate credence only by means of an unqualified appeal to conciliationism, an appeal that seems to simply ignore any skeptical worries raised by the disagreement about the validity of conciliationism. Conversely, if in response to the disagreement I do not adjust my credence for conciliationism, I will have been very non-conciliatory at the level of my final credences, but maximally conciliatory at the level of my reasoning about the disagreement, since I will have completely refrained from giving the disagreement the evidential force that my disputed conciliatory view would seem to require. And if I adopt some position between these two, then I will have been partly conciliatory and partly non-conciliatory both at the level of my final credences and at the level of my reasoning behind those credences. What is clear, then, is that to the extent that I respect the skeptical force of the disagreement at the level of my final credences, I will be disrespecting it at the level of my reasoning behind those credences, and vice versa.

It appears, then, that there is no principled way of respecting the skeptical pressures raised by a disagreement over conciliationism. For unlike most disagreements, which raise worries about the matter under dispute but not about how I should respond to the dispute, a disagreement over my belief in conciliationism simultaneously calls into question both my disputed belief and my views on how to respond to that dispute. We might say that in this case there are two vectors of skeptical force, both of the same magnitude but pointing in opposite directions. But of course the *net* force of mutually opposed vectors of equal magnitude is zero. Consequently, the disagreement ultimately has no evidential significance; or, more precisely, the concerns motivating conciliationism give me no reason for thinking that this disagreement has evidential significance. I am therefore free to base my credence for conciliationism purely on other evidential and rational factors.

This analysis suggests that there is a good reason for the conciliationist to exempt disagreements over conciliationism from the conciliatory prescriptions that apply to normal cases of disagreement. Moreover, this reason is the right type of "ground level reason" that addresses the concerns raised by the arbitrariness objection in a way that Elga's appeal to the high-level concern of consistency does not. For my account, unlike Elga's, is able to say just where someone errs who argues that because disagreements typically generate skeptical pressure

against one's disputed belief, disagreements over conciliationism also generate skeptical pressure against one's belief in conciliationism. On my analysis, this argument correctly identifies one vector of skeptical force, but it simply neglects to account for an additional and opposed vector of skeptical force that arises in cases where the disputed proposition is conciliationism (or some belief that is presupposed by conciliationism). The case for self-exemption can therefore be fully reconciled with the motivations for conciliationism, since the basis for self-exemption can be explained in terms of the dynamics of the disagreement-generated skeptical pressures posited by the conciliationist.

One might worry that my explanation for self-exemption only works when the conciliationist is in a disagreement with an anti-conciliationist and does not similarly work when someone disputes my conciliatory norm because they advocate some alternative conciliatory norm. For example, consider the following example, suggested by David Christensen as a challenge to my view.²² Suppose that I am a proponent of conciliationism, and I meet a redhead who is my apparent peer and who advocates CR, the view that conciliation is mandated only in disagreements with redheads. It may seem that in this case there are not two opposing vectors of skeptical force. For in this case, my redhead opponent calls into question my high credence for conciliationism but *endorses* the reasonability of my taking our disagreement as evidence against my conciliatory view. Thus, my giving evidential weight to the disagreement would not seem to be non-conciliatory at the level of my reasoning. So it would seem that all the skeptical pressures in this case point in the same direction.

But if we reflect a little further, I think it is clear that my conciliating in this case would be

²² The example was suggested in personal correspondence.

non-conciliatory at the level of my reasoning about the disagreement. For even though we both agree that, in general, one ought to conciliate in disputes with redhead peers, my only reason for thinking this is that I think that one ought to conciliate in disputes with all apparent peers, regardless of their hair color. Thus, if I were to conciliate in the dispute, I would have to rely on this more general reason, as I have no other basis for thinking conciliation is appropriate in this instance. And in relying on this general reason, I'd be relying on a belief that is incompatible with CR. So while my reduced credence for conciliationism would accord with CR, the reasoning behind that reduction in credence would still presuppose the falsity of CR and would thus be completely non-conciliatory. So the same point holds in this case: the pressure to reduce my credence in conciliationism is balanced by a pressure to *not* rely on the only rational basis I have for reducing my credence in conciliationism.

If I am right that the correct conciliatory norm will be self-exempting and that such self-exemption is not unacceptably arbitrary, then conciliationism is not epistemically self-defeating in contexts where it is disputed by apparent epistemic peers. With self-defeat concerns defused, we are free to assign conciliationism a credence that reflects our first-order assessment of its rational merits. And whatever the merits of conciliationism, the resolute conciliationist cannot be dismissed as dense or hypocritical. It is, rather, the weak-kneed variety that holds the less reasonable position.

3. THE CONCERN OF PROBLEMATIC GENERALIZATION

I have argued that there are good "ground level" reasons for thinking that the correct conciliatory policy will be self-exempting. The self-undermining problem therefore does not apply to the

most plausible version of conciliationism and need not worry the conciliationist. I now turn in this and the final section to two worries one might have about my position.

First, one might worry that my defense of self-exemption will generalize in problematic ways. It might seem that my defense of self-exemption, if translated to other contexts, would imply that certain self-defeat arguments that have been raised against other epistemic norms do not in fact constitute serious objections to those positions. And since in some cases it seems clear that these self-defeat arguments *do* constitute serious objections to the theories they target, one might conclude that my defense of self-exemption in the context of conciliationism should be rejected. Examples of self-defeat arguments that seem to pose legitimate challenges to the targeted theory might include Alvin Plantinga's (1981) arguments against "classical foundationalism" and George Bealer's (1992) arguments against empiricism. To this list we can add the example of method M in my elaboration of Elga's *Consumer Reports* example. Surely the fact that M ranks another ranking methodology more highly than itself points to a serious problem that cannot be justifiably dismissed by the move from M to self-exempting M*. If my argument for self-exemption in the context of conciliationism implies otherwise, then this would seem to be good reason for rejecting my position.

It will be helpful to more explicitly spell out the objection in the context of a specific example. Consider Bealer's self-defeat argument against empiricism. The version of empiricism that Bealer has in mind is Quine's, which, as characterized by Bealer, has three central dogmas:

(i) the principle of empiricism ("a person's experiences and/or observations comprise the person's *prima facie* evidence"); (ii) the principle of holism ("a theory is justified...for a person if and only if it is, or belongs to, the simplest comprehensive theory that explains all, or most, of

the person's *prima facie* evidence"); (iii) and the principle of naturalism ("the natural sciences [plus the logic and mathematics needed by them] constitute the simplest comprehensive theory that explains all, or most, of a person's experiences and/or observations") (99). According to Bealer, this version of empiricism is "epistemically self-defeating" since, by its own lights, empiricism is not justified (119). While I will not rehearse Bealer's detailed argument, the key claim is that the concept of epistemic justification (along with a host of other essential epistemological concepts) does not belong to the "simplest regimented formulation of the natural sciences," nor does any concept that is relevantly analogous to epistemic justification belong to that formulation (120 ff.). The natural sciences (and their supporting apparatuses) simply cannot say *anything* about epistemic justification. But if this is right, then it follows from the principles of empiricism and naturalism that the theory of empiricism (which employs these problematic epistemological concepts) does not belong to that theory which best explains a person's *prima facie* evidence. And from this and the principle of holism it follows that we are not justified in believing empiricism. So if empiricism is true, we are not justified in believing it.

It is clear that if Bealer's argument is sound, it presents a major challenge to those in the empiricist camp. For it seems that they will need to substantially revise their epistemology, perhaps to the point of altogether abandoning anything that could be reasonably labeled 'empiricist.' But now the objection to my defense of self-exempting concilationism can be rendered more explicit. For it might seem that if this defense is correct, then the challenge put forward by Bealer against the empiricist is not in fact a serious one. Here is how the objection against me might be articulated. Just as a disagreement over conciliationism generates two skeptical pressures that are mutually opposed and cancel one another out, so too empiricism's

lack of empirical support challenges both the empiricist's high credence for empiricism *and* her thinking that lack of empirical support is a sufficient reason to abandon a belief. We should therefore expect that the argument for self-exempting conciliationism would work here as well. If the argument for self-exempting conciliationism is correct, then the most rational theory in the empiricist camp is not empiricism, but is rather empiricism*, where empiricism* is like empiricism except that it is self-exempting and does not require belief in empiricism* to meet the same requirements as other beliefs in order to be justified.²³ This in turn implies that the self-defeat problem characterized by Bealer is not a serious problem for the empiricist, since it is not applicable to the most rational theory in the empiricist camp. But surely if Bealer's argument is right, the empiricist camp faces a major problem, one that is not easily solved by advocating empiricism* rather than empiricism. Since my account implies otherwise, it must be mistaken.

I readily concede that advocating empiricism* would be a completely unpersuasive and insufficient response to the challenge that Bealer raises against empiricism. But there is at least one critical disanalogy between the cases of conciliationism and empiricism that helps to account for why the move to self-exemption is acceptable in the first case but not in the second.²⁴

Whereas conciliationism is a theory characterizing a certain kind of second-order epistemic defeat, empiricism is a complete theory of first-order justification. This difference has important implications when we try to account for the first-order justification of the self-exempting

-

²³ Specifically, empiricism* modifies the "principle of holism" so that this principle applies to all beliefs except for the exempted class, which consists of empiricism* and any beliefs that must be presupposed if empiricism* is to be rationally believed. For beliefs in the exemption class, empiricism* holds that belonging to "the simplest comprehensive theory that explains all, or most, of the person's *prima facie* evidence" is *sufficient* for justification, but not *necessary* for justification.

²⁴ The following discussion benefited significantly from comments from [name removed].

versions of these two theories. To see this, consider first the question of empiricism*'s justification. Unlike empiricism, empiricism* is not a complete theory of justification: empiricism* gives necessary and sufficient conditions for justification that pertain to all beliefs except empiricism* and those beliefs that must be presupposed in order to rationally believe empiricism*. Empiricism* does offer us a universal sufficient condition for justification, one having to do with adequate basis in experience and observations; but if we accept Bealer's argument, we know that empiricism* itself does not satisfy this sufficient condition. So if empiricism* is justified, its justification must be explained by some theory of justification other than empiricism*, and this theory will have to account for the justification of empiricism* on non-empirical grounds. This means that if empiricism* is justified, then the correct complete theory of justification consists of empiricism* plus this other non-empiricist theory that accounts for the justification of empiricism*. But this means that believing empiricism* commits one to believing that there is a kind of non-empirical justification. So it turns out that the move to empiricism* is not a way of defending the basic empiricist position after all, but amounts to a departure from the essential commitments motivating empiricism. Moreover, we have good grounds for rejecting empiricism*. Empiricism* implies that all beliefs are justified on purely empirical grounds except for empiricism* (and any beliefs required for empiricism*'s justification that cannot be justified empirically), and clearly this gerrymandered theory is not plausible. For there is no reason to think that the scope of non-empirical justification should be limited to only this particular class of beliefs.

So Bealer's objection can be reframed as a dilemma for the empiricist: either the empiricist is not justified in his belief in unrestricted empiricism, or the empiricist endorses empiricism* and

is forced to admit a restricted role for non-empirical justification and to accept a theory of justification that is implausibly ad hoc. Given that neither option is acceptable, the empiricist ought to abandon his project and affirm a theory of justification that gives wider scope to non-empirical justification. And it seems that the same line of attack could be developed against any complete theory of justification (e.g., classical foundationalism) that fails to meet its own normative standards. In such cases, the move to a self-exempting version of the original self-defeating theory commits one to a picture of justification that is gerrymandered and implausible.

If conciliationism were a complete theory of first-order justification, then self-exempting conciliationism would be open to the same sort of objections I've just made against empiricism*. Suppose the conciliationist's view was that a belief is justified if and only if it is believed by the majority of experts. In this case, the move from unrestricted conciliationism to self-exempting conciliationism would raise questions about how the belief in self-exempting conciliationism is justified. And since the majority of experts do not accept self-exempting conciliationism. answering this question would require positing an implausibly gerrymandered picture of justification according to which all beliefs must be justified by expert consensus except for the belief in self-exempting conciliationism, which is justified in some other way. But of course conciliationism is not a complete theory of justification. Conciliationism is typically presented as a theory about a certain kind of second-order defeat. And even if conciliationism implies that the views of experts can play an important role in first-order justification, it does not imply that such views are the *only* source of justification. Thus, no special problems arise from the fact that the conciliationist must justify self-exempting conciliationism on grounds that have nothing to do with the views of experts. The conciliationist's larger theory of justification will already

acknowledge other sources of justification, so accounting for the justification of conciliationism will not require supplementing that theory in an ad hoc manner.

At last, I return to the example of M, Consumer Reports' imagined method for ranking products that, when applied to *Consumer Reports* and a competitor magazine using a different method, delivered the verdict that the competitor magazine does a better job at ranking products. As I already suggested, endorsing M*, a self-exempting version of M, would seem to be a very unreasonable response to the self-undermining of M. We are now in a position to see why M* is unreasonable even if self-exempting norms are reasonable in the cases of disagreement and Doubt Inducement. The case of M is more like the case of empiricism than that of conciliationism in that accounting for the justification of the self-exempting version of the theory requires one to subscribe to a (domain-specific) theory of justification that is implausibly gerrymandered. Since M supports the conclusion that Consume or Die is superior to Consumer Reports, in order to justify the judgment that Consumer Reports is in fact superior, one will presumably need to appeal to principles and standards of evaluation that are foreign to M. So if the editors of Consumer Reports respond to M's self-undermining by endorsing M* and continuing to stand by the superiority of their publication, then they will be committed to affirming an implausibly gerrymandered view on the correct methodology for ranking consumer products. For they will have to hold that all products should be ranked according to the principles put forward by M except for consumer product guides, which should be ranked by some other set of principles; and it is implausible to think that the principles that should be used to compare Consumer Reports and its competitors are principles that are applicable only to consumer product guides and not to products in any other categories. So if M self-undermines, the editors

of *Consumer Reports* ought to deny both M and M*, not because this is what M says it should do (a reason that would irrationally undercut itself) but because affirming M* commits them to unreasonable views concerning the correct ranking methodology.

4. MODERATING THE ACCOUNT

Resolute conciliationism, as I have characterized it thus far, would seem to license an absurd degree of intransigence in the face of overwhelming opposition. ²⁵ Suppose, for example, that Gary, a competent but unremarkable philosopher, begins to think about the epistemology of disagreement. Gary ends up endorsing a self-exempting conciliatory view, one that he calls RC (for "resolute conciliationism"). Gary then presents RC to the philosophical community, and scores of brilliant philosophers carefully consider its merits. While some think the view is on the right track and others think it is significantly mistaken, all of the philosophers independently conclude that one particular claim asserted by RC is incorrect. Let's call this claim P. Suppose that Gary hears all of his colleague's reasons for rejecting P, and though he acknowledges that the matter is complex, he does not ultimately find any of the reasons for rejecting P convincing. Ought Gary to nonetheless lower his credence for P in response to the fact that all of his philosophical colleagues reject it? The view I have been defending would seem to imply that Gary would be rational in remaining perfectly confident in the face of such overwhelming disagreement: given that conciliatory norms ought to be self-exempting, and that P is an implication of Gary's conciliatory norm RC, RC ought to exempt P from its conciliatory

²⁵ Christensen (2013) develops an objection of this sort in response to Elga's defense of self-exemption.

prescriptions. But surely it is absurd to think that Gary ought to be unmoved by such overwhelming consensus among his philosophical colleagues.

Any view that would prescribe a steadfast response in the case just described ought to be rejected as implausible. But in this final section, I will argue that the reasons I have given for self-exemption do not in fact support conciliatory positions that are completely immune from the skeptical threat of disagreement. The version of resolute conciliationism that properly emerges from the considerations developed above is, surprisingly, one that can accommodate the claim that Gary ought to conciliate in the above example.

The reason why conciliatory views should not be completely immune from being defeated by disagreement is that any real-life conciliationist will have multiple independent (or at least partially independent) conciliatory commitments, and while the arguments above show that an individual conciliatory commitment cannot rationally undermine itself, they do not rule out the rationality of one conciliatory commitment being my reason for taking disagreement to defeat another of my conciliatory commitments. To develop this point, consider the following toy example. Continuing with the case of Gary, suppose that Gary's conciliatory position consists of the following two claims:

COMMUNITY: In general, if I believe that p and then discover that a substantial portion of those who are best situated to assess p think that p is false, then I ought to significantly lower my credence for p.

INDIVIDUAL: In general, if the only knowledge I have about the distribution of opinion concerning p is that I believe that p and that one apparent epistemic peer with respect to p believes that p is false, then I ought to significantly lower my credence for p.

Of these two conciliatory commitments, INDIVIDUAL is obviously the more demanding, and also more questionable, commitment. So let us suppose that Gary is more confident in COMMUNITY than in INDIVIDUAL. Moreover, while Gary thinks that COMMUNITY has a significant chance of being true even if it turns out that INDIVIDUAL is false, he is certain that if COMMUNITY is false then INDIVIDUAL is also false. Given this setup, I suggest that the arguments I have given for self-exemption do *not* imply that Gary's belief in INDIVIDUAL ought to be completely immune from defeat by *community* disagreement. While the arguments for self-exemption imply that a commitment to INDIVIDUAL cannot give Gary a rational basis for doubting INDIVIDUAL, those arguments do not bar Gary from doubting INDIVIDUAL on the basis of his commitment COMMUNITY.

To see this, consider a case where it is known that most members of the community of experts on INDIVIDUAL (that is, those best positioned to assess that proposition's plausibility) think that INDIVIDUAL is false. How ought Gary to respond to this fact? The answer to this question will depend on whether INDIVIDUAL is contested for reasons that would also call into question COMMUNITY, or for reasons that leave COMMUNITY untouched, or for both sorts of reasons.

Consider first the scenario where INDIVIDUAL is contested for reasons that would equally call COMMUNITY into question. On any reasonable conciliatory approach, conciliatory requirements

will apply not only at the level of conclusions, but also at the level of reasons for those conclusions. ²⁶ Thus, even if the disagreement is ostensibly over the plausibility of INDIVIDUAL, the disagreement will pose just as much challenge to Gary's belief in COMMUNITY. But this means that the disagreement does not in fact give Gary a reason for conciliating. For to the extent that the disagreement calls into question Gary's high credence for COMMUNITY, it *also* calls into question the reasonability of his taking expert disagreement as a reason for lowering his confidence. In this case, then, the pressure to adopt a lower credence is counterbalanced by the pressure to *not* rely on COMMUNITY, which is Gary's only basis for conciliating. INDIVIDUAL is thus protected from defeat by community disagreement if that disagreement equally calls into question COMMUNITY.

Next, consider the scenario where INDIVIDUAL is contested for reasons that would *not* call COMMUNITY into question. In this case, Gary can increase his confidence in his disputants' reasons for contesting INDIVIDUAL *without* thereby gaining reasons for doubting COMMUNITY. Thus, COMMUNITY can serve as a rational basis for Gary's conciliating with respect to INDIVIDUAL. Finally, if INDIVIDUAL is contested both for reasons that also implicate COMMUNITY and for reasons that do not implicate COMMUNITY, COMMUNITY can give Gary a reason for being conciliatory with respect to the latter but not to the former group of reasons.

The key point is that if a conciliationist has multiple independent (or partially independent) conciliatory commitments, the arguments for self-exemption do *not* license packaging all of these commitments into one conciliatory norm and then exempting all aspects of the norm from

_

For if I conciliate in a disagreement over p but not with respect to my disputant's reasoning about p, I'll end up in an inconsistent position, where my credences for reasons bearing on p are inconsistent with my credence for p.

conciliatory prescriptions. To the extent that two conciliatory commitments are noetically independent, it may be perfectly possible for one of these commitments to rationally ground skepticism towards the other commitment when this other commitment is disputed. As long as the conciliatory commitment that is motivating the doubt is not also implicated by the dispute, then there will not be mutually opposed vectors of skeptical force; moreover, conciliation in such a case would not be open to the no justification objection or the conditional credence objection. So in Gary's case, COMMUNITY and INDIVIDUAL essentially constitute two distinct conciliatory norms, each applying to different sorts of disagreements. While fully-articulated versions of these norms ought to be self-exempting, such self-exemption would not automatically rule out the possibility of one norm serving as basis for doubting the other norm. We could alternatively speak of Gary's conciliatory "norm" in the singular, referring to the entire package of his conciliatory views. But since such an encompassing norm will be the product of several conciliatory commitments that are at least partially independent, this norm should not exempt itself in its entirety from conciliatory prescriptions. Instead, it should allow for some of its components to be undermined by disagreement as long as this disagreement does not also implicate the particular conciliatory commitment that rationally grounds the conciliation.

The picture that emerges from this discussion is one where my total conciliatory view consists of a nest of conciliatory commitments, some tame and highly plausible and some demanding and less plausible. None of these commitments can rationally undermine themselves, though this does not prevent them from being undermined by my other conciliatory commitments. My more questionable conciliatory commitments with more demanding requirements will be especially susceptible to being defeated by disagreement, since such

commitments will likely be contested partly for reasons that do not apply to my tamer conciliatory commitments, making it possible for these tamer commitments to undermine my belief in the more demanding requirements. But such defeat would not be an instance of selfdefeat, since the basis for the defeat would be supplied by some other commitment. Of course the tamest and most general of my conciliatory commitments (which we might characterize as follows: disagreement should sometimes lead me to reduce my confidence in my disputed belief) will be completely immune from defeat by disagreement, since none of my more specific conciliatory commitments could lead me to doubt this commitment without also undercutting their own justification. But even if the entire world were to dispute this most tame of conciliatory commitments, my sticking by the commitment would not be objectionably immodest. For bringing myself to conciliate in this case would require that I rely on conciliatory reasoning. reasoning that is universally disputed. So whether I conciliate or not, I will find myself endorsing (whether explicitly in my conclusions or implicitly in the reasoning behind my conclusions) a way of thinking that I know to be universally disputed. Epistemic modesty would simply not be an option. And where modesty is impossible, immodesty is not objectionable.

Where does my account leave today's conciliationist? Can she, in the face of disagreement, remain resolute in her controversial conciliatory views? Or should she be led by her tamer and less controversial "conciliatory" commitments to back down from any position with significant conciliatory bite? This is no easy question, depending as it does on a number of empirical questions concerning the various commitments of a given conciliationist and the kinds of reasons motivating the opponents of various conciliatory proposals. But without considering specific conciliatory proposals, we can arrive at some general conclusions about what kind of

conciliatory positions are likely to be viable in a context like our own where conciliationism is contested. First, a detailed conciliatory norm that incorporates specific provisions that are controversial even among conciliationists is likely to be defeated by disagreement. For however confident one may initially be that some such norm is correct, one is likely to be even more firmly committed to a more vaguely-stated (and thus less controversial) conciliatory norm that is not subject to all of the criticisms directed toward the detailed norm and that supports the conclusion that one ought to conciliate in the disagreement over the merits of the more detailed norm.

Matters are quite different, however, when we consider those vaguer conciliatory commitments that serve to mark the boundary between conciliationists and non-conciliationists. Let *NI* be some vague but nonetheless fairly demanding conciliatory norm, one that would be rejected by almost all opponents of conciliationism and accepted by almost all of its supporters. (There are several different ways we could imagine characterizing *NI*. Perhaps *NI* could offer some paradigm examples where significant conciliation is required and then assert that in disagreements relevantly like one of the paradigms, significant conciliation is required. Or perhaps *NI* could point to some of the principal conciliatory proposals in the literature and then assert that these are reasonably good approximations of the truth.²⁷) One consequence of the vagueness of *NI* is that those who dispute it are likely to have relatively little to disagree with beyond some very general commitments that are arguably quite fundamental to any position deserving of the label 'conciliationism.'

.

²⁷ For an example of a conciliatory commitment characterized along these lines, see Frances (2010, 459).

It is probable, then, that NI would not be defeated by the disagreement over the view. For NI would be defeated by disagreement only if there is some other norm N0 that (i) is affirmed by advocates of NI; (ii) delivers a conciliatory prescription when applied to the type of opposition faced by NI; and (iii) would not itself be undermined were one to accept the reasons that opponents of N1 use to attack N1. Plausibly, there is no N0 that satisfies all of these criteria. For given the current state of the dispute over conciliationism, (ii) would seem to demand that N0 be a rather demanding requirement (after all, it is not as though the overwhelming majority of experts have rejected conciliationism), while (iii) seems to require that N0 be a very weak requirement (since if it is a demanding requirement, the criticisms lodged against the very general NI will probably apply to N0 as well). I suspect that if one asked opponents of conciliationism whether reasonability would require that they give up some philosophical position of theirs if it were to face the same degree and type of opposition faced by conciliationism, they would answer no, adding that any norm that required them to conciliate in such a situation would fall afoul of the same critiques that they advance against conciliationism. If I'm right in thinking that they would answer this way, then there is good reason to think that there is no norm that satisfies (i) – (iii), and that a conciliationist who accepts NI may reasonably stand by N1 even in the face of disagreement.

So even if the reasonable conciliationist cannot be *entirely* resolute in the face of disagreement over conciliationism, the conciliationist need not be entirely conciliatory either. Belief in highly specific conciliatory norms will probably have to be given up if the specifics of such norms are contested for reasons that do not simultaneously call into question more general conciliatory commitments. But it seems probable that one can reasonably maintain belief in these

more general conciliatory commitments even in the face of the current disagreement over the merits of conciliationism.²⁸

BIBLIOGRAPHY

- Bealer, George. 1992. "The Incoherence of Empiricism." *Proceedings of the Aristotelian Society, Supplementary Volumes* 66: 99–143.
- Bergmann, Michael. "Rational Disagreement after Full Disclosure." *Episteme* 6, no. 3 (October 2009): 336–353.
- Christensen, David. 2007. "Epistemology of Disagreement: The Good News." *Philosophical Review* 116 (2): 187–217.
- ——. 2011. "Disagreement, Question-Begging and Epistemic Self-Criticism." *Philosopher's Imprint* 11 (6) (March): 1–22.
- ——. 2013. "Epistemic Modesty Defended." In *Disagreement: New Essays*, edited by David Christensen and Jennifer Lackey, 77–97. Oxford: Oxford University Press.
- Elga, Adam. 2007. "Reflection and Disagreement." Noûs 41 (3): 478-502.
- ——. 2010. "How to Disagree About How to Disagree." In *Disagreement*, edited by Richard Feldman and Ted Warfield, 175–186. New York: Oxford University Press.
- Feldman, Richard. 2006. "Epistemological Puzzles About Disagreement." In *Epistemology Futures*, edited by Stephen Hetherington, 216–236. New York: Oxford University Press.
- Field, Hartry. 2000. "Apriority as an Evaluative Notion." In *New Essays on the A Priori*, edited by Paul Boghossian and Christopher Peacocke, 1:117–150. New York: Oxford University Press.
- Frances, Bryan. 2010. "The Reflective Epistemic Renegade." *Philosophy and Phenomenological Research* 81 (2): 419–463.
- Gutting, Gary. 1982. *Religious Belief and Religious Skepticism*. Notre Dame, IN: University of Notre Dame Press.
- Kelly, Thomas. 2010. "Peer Disagreement and Higher Order Evidence." In *Disagreement*, edited by Richard Feldman and Ted A. Warfield, 111–174. New York: Oxford University Press.

55

²⁸ For helpful discussions of various versions of this paper, I am grateful to Keith DeRose, Lara Buchak, Dustin Locke, Daniel Immerman, Alex Worsnip, David Christensen, Sun-Joo Shin, Michael Della Rocca, three anonymous referees, participants in two seminars at Yale, and participants in the 2011 Notre Dame and Northwestern Graduate Student Epistemology Conference.

- Kornblith, Hilary. 2010. "Belief in the Face of Controversy." In *Disagreement*, edited by Ted Warfield and Richard Feldman, 29–52. New York: Oxford University Press.
- Kvanvig, Jonathan L. 2011. "The Rational Significance of Reflective Ascent." In *Evidentialism* and *Its Discontents*, edited by Trent Dougherty, 34–54. New York: Oxford University Press.
- Lackey, Jennifer. 2010. "What Should We Do When We Disagree?" Oxford Studies in Epistemology 2: 274–93.
- Lewis, David. 1971. "Immodest Inductive Methods." Philosophy of Science 38 (1): 54-63.
- Plantinga, Alvin. 1981. "Is Belief in God Properly Basic?" Noûs 15 (1): 41-51.
- Thune, Michael. 2010. "'Partial Defeaters' and the Epistemology of Disagreement." *The Philosophical Quarterly* 60 (239) (April): 355–372.
- Weatherson, Brian. 2007. "Disagreeing About Disagreement". Unpublished manuscript. http://brian.weatherson.org/DAD.pdf.
- ———. 2010. "Do Judgments Screen Evidence?" Unpublished manuscript. http://brian.weatherson.org/JSE.pdf.
- Weiner, Matt. 2007. "More on the Self-undermining Argument." *Opiniatrety*. http://mattweiner.net/blog/archives/000781.html.