

AN ARROVIAN IMPOSSIBILITY THEOREM FOR THE EPISTEMOLOGY OF DISAGREEMENT

Nicholaos JONES

ABSTRACT: According to conciliatory views about the epistemology of disagreement, when epistemic peers have conflicting doxastic attitudes toward a proposition and fully disclose to one another the reasons for their attitudes toward that proposition (and neither has independent reason to believe the other to be mistaken), each peer should always change his attitude toward that proposition to one that is closer to the attitudes of those peers with which there is disagreement. According to pure higher-order evidence views, higher-order evidence for a proposition always suffices to determine the proper rational response to disagreement about that proposition within a group of epistemic peers. Using an analogue of Arrow's Impossibility Theorem, I shall argue that no conciliatory and pure higher-order evidence view about the epistemology of disagreement can provide a true and general answer to the question of what disagreeing epistemic peers should do after fully disclosing to each other the (first-order) reasons for their conflicting doxastic attitudes.

KEYWORDS: Arrow's theorem, epistemology, higher-order evidence, peer disagreement, rationality

1. Views about the Epistemology of Disagreement

Situations often arise in which we find ourselves disagreeing with our peers. Even when we have access to the same evidence and respond to that evidence in equally reliable ways, we sometimes form conflicting beliefs. This occurs, for example, when jurors reach different judgments about a defendant's guilt; when meteorologists offer competing weather forecasts; when philosophers do metaphysics; when scientists offer conflicting accounts of experimental data; when physicians pronounce different causes for the same diseases; when politicians make different policy recommendations for addressing social issues; and so on.¹

¹ Richard Feldman, "Epistemological Puzzles about Disagreement," in *Epistemology Futures*, ed. Stephen Hetherington (Oxford: Oxford University Press, 2006), 216-236; Thomas Kelly, "Peer Disagreement and Higher-Order Evidence," in *Disagreement*, eds. Richard Feldman and Ted Warfield (Oxford: Oxford University Press, 2010), 111-174.

There are several competing views about the proper rational response to disagreement within a group of epistemic peers. According to *conciliatory* views, when epistemic peers have conflicting doxastic attitudes toward a proposition and fully disclose to one another the reasons for their attitudes toward that proposition (and neither has independent reason to believe the other to be mistaken), each peer should always change his attitude toward that proposition to one that is closer to the attitudes of those peers with which there is disagreement.² *Steadfast* views, in contrast, maintain that when epistemic peers have conflicting doxastic attitudes toward a proposition and fully disclose to one another the reasons for their attitudes toward that proposition, sometimes some peers may maintain their original attitude toward the proposition.³

There are also competing views about the kind of evidence that determines the proper rational response to disagreement within a group of epistemic peers. *First-order evidence* for a proposition is any evidence that bears directly on that proposition's truth-value. First-order evidence can include perceptual evidence, testimonial evidence, inferential evidence, intuition, and so on. *Higher-order evidence* for a proposition, in contrast, is evidence about first-order evidence for that proposition.⁴ For example, a person's higher-order evidence for the proposition that God exists might include the fact that a peer takes the ontological argument to be sound, the fact that another peer takes the evidential problem of evil to conclusively refute God's existence, the fact that a peer takes reports of personal experience to be evidence for God's existence, and so on. Higher-order evidence is evidence about what first-order evidence supports.

There is disagreement about whether higher-order evidence for a proposition always suffices to determine the proper rational response to disagreement about that proposition within a group of epistemic peers. According to what I shall call *pure higher-order evidence* (HOE) views, it does. For example, according to the equal weight view, when two peers adopt conflicting doxastic attitudes toward a proposition after full disclosure, the rational response to that disagreement depends upon what those attitudes are and nothing more. *Mixed evidence* views, in contrast, maintain that sometimes first-order evidence about a proposition helps to

² See David Christensen, "Epistemology of Disagreement: The Good News," *Philosophical Review* 116 (2007): 187-217; Adam Elga, "Reflection and Disagreement," *Noûs* 41 (2007): 478-502.

³ I take this terminology from David Christensen, "Disagreement as Evidence: The Epistemology of Controversy," *Philosophy Compass* 4/5 (2009): 756. See also Kelly, "Peer Disagreement and Higher-Order Evidence" and Thomas Kelly, "The Epistemic Significance of Disagreement," in *Oxford Studies in Epistemology*, Volume I, eds. Tamar Szabò Gendler and John Hawthorne (Oxford: Oxford University Press, 2005), 167-196.

⁴ See Kelly, "The Epistemic Significance of Disagreement," 185-190.

determine the proper rational response to disagreement. Examples of what I am calling mixed views are Kelly's total evidence view and Lackey's justificationist view.⁵

I shall argue that no view that is both conciliatory and pure HOE can provide a true and general answer to the question of what disagreeing epistemic peers should do after fully disclosing to each other the (first-order) reasons for their conflicting doxastic attitudes. As a matter of principle, any such view is committed to two constraints about the way in which the rational response to disagreement among epistemic peers is a function of those peers' higher-order evidence. These constraints, and an additional adequacy condition for all views of peer disagreement, are formal analogues to the ones that appear in Arrow's Impossibility Theorem.⁶ This analogy, together with replies to potential objections, show that conciliatory pure HOE views are either false or unacceptably ad hoc.

I begin, in the first section, with some preliminaries about how to understand the formal structure of peer disagreement situations in a way that makes Arrow's Theorem relevant to the epistemological debate. Next, I motivate an adequacy condition for views about peer disagreement. Then I argue that conciliatory pure HOE views are committed to two additional constraints about the way in which the rational response to disagreement among epistemic peers is a function of those peers' higher-order evidence. After presenting a formal analogue of Arrow's Impossibility Theorem, I consider some ways in which conciliatory pure HOE views might attempt to avoid the upshot of Arrow's Impossibility Theorem as applied to the epistemology of disagreement. I conclude that such views are false if as they cannot avoid the theorem, and unacceptably ad hoc if they can.

2. Abstract Structure of Peer Disagreement Situations

There are at least three doxastic attitudes possible toward any proposition. These attitudes might be course-grained: believing the proposition; disbelieving it (believing it is false); and withholding judgment about it (neither believing nor disbelieving it).⁷ They might be fine-grained, such as attitudes that involve

⁵ In Kelly, "Peer Disagreement and Higher-Order Evidence" and Jennifer Lackey, "What Should We Do When We Disagree?" in *Oxford Studies in Epistemology*, Volume 3, eds. Tamar Szabò Gendler and John Hawthorne (Oxford: Oxford University Press, 2010), 274-293.

⁶ See Kenneth J. Arrow, "A Difficulty in the Concept of Social Welfare," *The Journal of Political Economy* 58 (1950): 328-346; David Austen-Smith and Jeffrey S. Banks, *Positive Political Theory I: Collective Preference* (University of Michigan Press, 1999); John Geanakoplos, "Three Brief Proofs of Arrow's Impossibility Theorem," *Economic Theory* 26 (2005): 211-215.

⁷ See Jane Friedman, "Suspended Judgment," *Philosophical Studies* (forthcoming). DOI: 10.1007/s11098-011-9753-y.

confidence levels: believing the proposition with degree of confidence 1; believing it with degree of confidence 0.5; and so on. Kelly argues that certain conciliatory views should adopt a fine-grained analysis of the possible doxastic attitudes.⁸ Nothing I say below depends upon whether there are exactly three possible doxastic attitudes, as a course-grained approach suggests, or more than three, as a fine-grained approach suggests. My argument requires only that there are at least three such attitudes, whatever they happen to be.

The literature on the epistemology of disagreement considers situations in which epistemic peers adopt differing doxastic attitudes toward a proposition after fully disclosing to each other the reasons for their attitude toward that proposition. Christensen defines two people as *epistemic peers* regarding a proposition just in case they have considered roughly the same evidence with respect to whether that proposition is true and they are roughly equally good at responding to that kind of evidence.⁹ While there are other definitions available in the literature, this suffices as a working definition. Nothing in my argument hinges upon its correctness. I require only that there are at least two epistemic peers. When there are not at least two peers, my argument does not hold. But since the situations of interest to epistemologists are those in which there is disagreement, and since every disagreement involves at least two peers, this limitation is not significant.

Moreover, all of the situations of interest to epistemologists are ones in which epistemic peers adopt different doxastic attitudes toward the same proposition. I shall say that an attitude a person adopts toward a proposition is ON for that person with respect to that proposition, and that an attitude a person does not adopt toward a proposition is OFF. For example, if the possible doxastic attitudes are coarse-grained and if, regarding the proposition that God exists, believing is the only attitude Aquinas has toward it, then *believing that God exists* is ON for Aquinas while both *disbelieving that God exists* and *withholding judgment about God's existence* are OFF. There is peer disagreement regarding a proposition when the peers have different doxastic attitudes ON toward that proposition.

Regardless of what the possible doxastic attitudes are, each of a person's possible doxastic attitudes toward a proposition is either ON or OFF for that person toward that proposition. But it seems that there are situations in which one attitude can be *less* OFF (or more ON) for a person toward a proposition than another. For example, imagine a theist and atheist discussing whether God exists. Suppose the conversation turns to agnosticism, the view that our available evidence does not warrant either believing or disbelieving that God exists. Further suppose that they

⁸ Kelly, "Peer Disagreement and Higher-Order Evidence," 117-118.

⁹ Christensen, "Epistemology of Disagreement," 211.

both admit that agnosticism is more plausible than their opponent's view, even though each retains her belief. Then *withholding judgment about God's existence* is less OFF for the atheist than is *believing that God exists*, and it is less OFF for the theist than is *disbelieving that God exists*. Surely this kind of situation is common; but it is a situation in which people's rankings of possible doxastic attitudes have more than two levels. Also consider

Ranking. Lucy is on *Let's Make a Deal*. Lucy will only choose a door when she believes the prize is behind it; otherwise, she will walk away from the game rather than make a choice. Lucy initially believes that the prize is behind the leftmost of three doors, and so she chooses that door. Regardless of which door Monty Hall reveals to contain a goat, Lucy will continue to believe that there is a prize behind one of the two unopened doors, and in fact she will come to believe that the prize is behind the unchosen and unopened door. She will not walk away from the game.

Let the proposition R be: *The prize is behind the rightmost of the three doors*. When Lucy initially chooses the leftmost door, the attitude *disbelieving R* is ON for her, while the attitudes *believing R* and *withholding judgment about R* are OFF. But it seems that, prior to Monty Hall opening one of the two unchosen doors, *believing R* is *less OFF* for Lucy than *withholding judgment about R*. For, at that time, she is more disposed to change from disbelieving R to believing R than she is to change from disbelieving R to withholding judgment about R. When Lucy is disbelieving R, the (non-actual) possible world in which she believes R is closer than the world in which she withholds assent about R. Given this, it seems that when *disbelieving R* is ON for Lucy, *believing R* is less OFF for her than is *withholding judgment about R*. (For similar reasons, it seems that, when Lucy chooses the leftmost door, *believing that the prize is behind the center door* is less OFF for her than is *withholding judgment about whether the prize is behind the center door*.)

The *is less OFF than* relation is obviously *transitive*: for any person S, proposition P, and distinct doxastic attitudes X,Y,Z toward P, whenever X is less OFF for S than is Y and Y is less OFF for S than is Z, X is less OFF for S than is Z. Regarding the Ranking case, transitivity entails that, when Lucy initially chooses the leftmost door, *disbelieving R* is less OFF for her than is *withholding judgment about R*, because any ON attitude is less OFF than any OFF attitude.

Transitivity is an essential presupposition for the Arrovian-style impossibility theorem for conciliatory pure HOE views of peer disagreement. Also essential is a modal claim about rankings of doxastic attitudes toward propositions.

Nicholaos Jones

Depth: It is possible that there exists a person S, proposition P, and distinct doxastic attitudes X,Y,Z toward P such that X is less OFF for S than is Y and Y is less OFF for S than is Z.

The Ranking case supports *Depth*. When Lucy initially chooses the leftmost door, *disbelieving R* is less OFF for her than is *believing R* (by virtue of *disbelieving R* being ON) and *believing R* is less OFF for her than is *withholding judgment about R*. *Depth* entails that the *is less OFF than* relation orders people's doxastic attitudes toward propositions in a way that does not necessarily have only two ranking levels.

3. Response Functions and Doxastic Attitude Rankings

Pure HOE views about peer disagreement may be understood as maintaining that there is a function that takes as input information about higher-order evidence about disagreeing peers' doxastic attitudes toward a disputed proposition and yields as output a verdict about the rational response to that disagreement after the peers disclose to each other the (first-order) reasons for their conflicting attitudes. For example, the equal weight view may be understood as maintaining that the following function is correct for the case in which two epistemic peers disagree about some proposition P

$$(EWV): (C_1 + C_2)/2 = C_R,$$

where C_1 is the credence peer 1 gives to P, C_2 is the credence peer 2 gives to P, and C_R is the credence each peer ought to give to P after full disclosure.¹⁰ Similarly, the extra weight view may be understood as proposing as correct the function

$$(XWV): (C_1+C_2)/2 + x(C_1-C_2)/2 = C_R,$$

where peer 1 is (indexically) the person adjusting her doxastic attitude and x ($0 \leq x \leq 1$) is the amount of extra weight that peer gives to her attitude.

Let us call functions like EWV and XWV *response functions* and information about a peer's doxastic attitudes toward a proposition a *doxastic profile* for that peer. Then pure HOE views may be understood as maintaining that the rational response to peer disagreement is determined by a response function that takes as input the doxastic profiles for all disagreeing peers and yields as output a doxastic profile that those peers ought to have after full disclosure. Conciliatory views may

¹⁰ For an objection and alternative to this way of understanding the equal weight view, see Branden Fitelson and David Jehle, "What is the Equal Weight View?" *Episteme* 6 (2009): 280-293.

be understood as adding that the output of this response function should be some kind of compromise among the profiles taken as input.

Response functions need not be mathematical. Consider, for example, Feldman's split the difference view.¹¹ According to this view, if one peer believes P and another peer disbelieves P, the rational response to this disagreement after these peers disclose their reasons to each other is for each peer to withhold assent about P. This may be represented as a non-mathematical function f_r , where $B_n(P)$ represents that peer n believes that P:

$$(SDV): f_r(B_1(P), B_2(\neg P)) = \neg B(P) \& \neg B(\neg P).$$

While SDV itself has the appearance of a mathematical equation, the function f_r is not mathematical, in the same way that the function $f_{\&}(P, Q)$ for conjunction-introduction is not mathematical.

The output to a response function need not be a doxastic profile in which there is a *unique* doxastic attitude that disagreeing peers ought to have after full disclosure. Some pure HOE views, like the equal weight view, maintain that there is exactly one doxastic profile all peers ought to have after full disclosure; others, like the extra weight view, allow peers to have different profiles after full disclosure by virtue of advocating *indexical* response functions. There even could be non-indexical response functions that allow more than one doxastic attitude as the rational response to peer disagreement after full disclosure.¹² Accordingly, understanding pure HOE views in terms of response functions is neutral regarding whether, for any given evidential situation, there is only one rational response to peer disagreement after full disclosure in that situation.¹³

Information about the doxastic profiles taken as input for conciliatory pure HOE response functions cannot be *merely* information about which doxastic attitudes happen to be ON for the peers, even though typical presentations of such views give this impression. For there is some reason to think that, if the input were restricted in this way, conciliatory pure HOE views would face insuperable difficulties.

Consider a situation in which two epistemic peers, an atheist and an agnostic, are discussing whether God exists. Suppose that there are three possible doxastic attitudes: believing; disbelieving; withholding assent. Conciliatory views about disagreement entail that, after full disclosure, each peer should change his doxastic

¹¹ Feldman, "Epistemological Puzzles about Disagreement."

¹² For some suggestions, see Kelly, "Peer Disagreement and Higher-Order Evidence," 120-121.

¹³ For further discussion of uniqueness, see Roger White, "Epistemic Permissiveness," *Philosophical Perspectives* 19 (2005): 445-459.

attitude in the direction of the other. But, as Kelly notes, there is no suitable way to do so.¹⁴ Kelly takes this to entail that conciliatory views should adopt a more fine-grained approach to possible doxastic attitudes. But this precludes the problem only if those attitudes are dense, so that there is always another attitude between any distinct doxastic attitudes. For if the attitudes are discrete, conciliatory views flounder in situations where disagreeing peers adopt conflicting attitudes toward a proposition and there is no "middle-ground" attitude available. However, it is extremely unlikely that the doxastic attitudes had by actual people are just as fine-grained as, say, the real numbers. So Kelly's proposal on behalf of conciliatory views preserves their truth at the cost of rendering them inapplicable to the actual world.

Conciliatory views about disagreement can avoid the preceding difficulty without endorsing an unrealistic view about possible doxastic attitudes, by allowing input to response functions to include more than information about which attitudes happen to be ON for the peers after full disclosure. For conciliatory views that are also pure HOE views, this further information must be information about higher-order evidence. The only such information is information about how peers rank possible doxastic attitudes in terms of the *is less OFF than* relation. Fortunately, this solves the problem without the costs of Kelly's proposal.

Consider again the disagreeing atheist and agnostic. The atheist has *disbelieving that God exists* ON, while the agnostic has *withholding judgment about whether God exists* ON. Since their doxastic attitudes differ, the rest of their doxastic profiles must differ as well. For example, perhaps the atheist's profile is such that: *disbelieving that God exists* is less OFF than both *withholding assent that God exists* and *believing that God exists*, while neither of these latter two attitudes is less OFF than the other; and perhaps the agnostic's profile is such that *withholding assent that God exists* is less OFF than both *disbelieving that God exists* and *believing that God exists*, while neither of these latter two attitudes is less OFF than the other. If conciliatory views require only that two disagreeing peers change their doxastic *profiles* toward each other (rather than change the *attitudes* that they happen to have ON) after full disclosure, such views can maintain that the disagreeing peers should change their rankings of attitudes that are OFF. So, for example, in the case of the atheist and agnostic, such a view might maintain that the atheist should adopt a profile in which *disbelieving that God exists* is less OFF than *withholding assent that God exists*, which in turn is less OFF than *believing that God exists*, and that the agnostic should adopt one in which *withholding assent that God exists* is less OFF than *disbelieving that God exists*, which in turn is less OFF than *believing that God exists*. This kind of response to peer disagreement does

¹⁴ Kelly, "Peer Disagreement and Higher-Order Evidence," 117.

not *remove* the disagreement between the atheist and the agnostic; but then, other conciliatory pure HOE views, such as the extra weight view, also allow the disagreement to persist. Since disagreeing peers are guaranteed to have differing doxastic profiles, *some* kind of change among the OFF attitudes for each peer is always possible. Accordingly, conciliatory pure HOE views can avoid the problem Kelly raises without making themselves inapplicable to the actual world, provided that they propose response functions that take as input information about the rankings in peers' doxastic profiles.

Extant conciliatory and pure HOE views of peer disagreement do not consider response functions that take this kind of information as input. Nor, for that matter, do steadfast or mixed evidence views. For this reason, the peer disagreement literature has yet to consider adequacy conditions for such response functions. One *prima-facie* plausible condition is that, for any pair of distinct doxastic attitudes, such functions should yield as output a relative ranking of those attitudes that is independent of changes in peers' doxastic profile rankings for other pairs of attitudes after the peers fully disclose to each other the reasons for their attitudes.

IIA: For any proposition P and any distinct doxastic attitudes X,Y toward P, if some or all peers change their doxastic profiles toward P after full disclosure without changing the relative ranking of X and Y within those profiles, the output of the response function does not change the relative ranking of X and Y.

(*IIA* abbreviates *I*ndependence of *I*rrelevant *A*lternatives.) Consider an abstract situation in which, for some proposition P and doxastic attitudes X,Y, and Z toward P, the output of the response function yields that X should be less OFF than both Y and Z. This output is based upon full disclosure of all evidence among epistemic peers and, perhaps, the doxastic profiles of the peers after this disclosure. The output is either eternally correct for the peers' evidential situation or not. If it is eternally correct, then if some of the peers change their doxastic profiles without acquiring new evidence (or losing available evidence), the output should remain as it was initially, because the peers' evidential situation remains the same. This accords with *IIA*. If the initial output is not eternally correct, the updated output of the response function depends, at least in part, upon the changed doxastic profiles of the peers. The intuition driving *IIA* in this condition is that updates to response function output should be proportionate to changes in peers' doxastic profiles. (If a peer changes the relative ranking of attitudes X and Y but not the relative ranking of X and Z, then if the response function output requires updating, I say that the updating is *proportionate* just if the function's output changes the relative ranking of attitudes X and Y but not the relative ranking of X and Z.) The motivation for this intuition is that, when a peer changes one pairwise ranking of doxastic

attitudes but not other pairwise rankings despite acquiring no new evidence (and losing no available evidence), there is no reason that warrants changing any of the other pairwise rankings, because all the initial evidence is the same; and when nothing warrants a change in pairwise rankings that are rational, changing those rankings would be irrational. If, say, there is no reason that warrants changing the rational relative ranking of attitudes X and Z, changing this relative ranking would be irrational, and so the response function's updated output regarding the relative ranking of X and Z should remain unchanged.

4. Constraints on Conciliatory Pure HOE Response Functions

Conciliatory pure HOE views impose two conditions on response functions that make them incompatible with *IIA*. The first is that there is no peer such that that peer's ranking one doxastic attitude as less OFF toward a proposition than another after full disclosure strictly implies that output of the response function ranks the former attitude as less OFF toward that proposition than the latter attitude.

Fallibility: It is not the case that there exists a peer such that, for any proposition P and any distinct doxastic attitudes X,Y toward P, necessarily, whenever that peer ranks X as less OFF than Y after full disclosure, the response function yields as output a ranking in which X is less OFF than Y.

All conciliatory views about disagreement endorse *Fallibility*. If *Fallibility* were false, then there could be a peer disagreement in which at least one party to the dispute is not required, after full disclosure, to change his attitude toward the disputed proposition to one that is closer to the attitudes of those peers with which he disagrees. But, according to conciliatory views, such change is always required of all peers.

The second condition on response functions for conciliatory pure HOE views concerns situations in which all peers have the same pairwise ranking of distinct possible doxastic attitudes toward a proposition after full disclosure.

Unanimity: For any proposition P and any distinct doxastic attitudes X,Y toward P, if all peers rank X as less OFF toward P than Y after full disclosure, the response function yields as output a ranking in which X is less OFF than Y.

For example, according to *Unanimity*, if everyone flat-out believes that the continuum hypothesis is true after fully disclosing to each other the reasons for their belief, the rational response to this situation is to rank *believing the continuum hypothesis* as less OFF than both *disbelieving the continuum hypothesis* and *withholding judgment about the continuum hypothesis*. If *Unanimity* is false, then there is some proposition P and distinct attitudes X,Y such that, although all

peers rank X as less OFF toward P than Y after full disclosure, those peers ought to change their doxastic profiles so as to not rank X as less OFF toward P than Y. However, according to pure HOE views, no peer in such a situation has any evidence to support changing her original assessment of the evidence for P, and so no peer ought to change her original doxastic profile after full disclosure.

All pure HOE views endorse *Unanimity*. For example, according to both the equal weight view and the extra weight view, if everyone has a credence of 0.9 toward P after full disclosure, having a credence of 0.9 toward P is the rational attitude to have. (Strictly speaking, pure HOE views do not apply to cases of unanimous peer *agreement*; but they should extend naturally to such cases in a way that validates *Unanimity*.) The falsity of *Unanimity* opens the possibility that, even if everyone has a credence of 0.9 toward P after full disclosure, that is not the rational credence to have, because some other credence should be less OFF toward P. But if everyone's evidence leads them to have a credence of 0.9 toward P after full disclosure, no one has reason to revise their credence. Also, consider

Ranking 2. Before Monty Hall opens the center door for Lucy, Lucy consults Marilyn, her off-stage friend. Lucy discovers that Marilyn also believes that the prize is behind the leftmost door, that Marilyn will continue to believe that the prize is behind *some* door no matter which one Monty opens, and that Marilyn will come to believe that the prize is behind the unchosen and unopened door after Monty opens a door.

Before Lucy consults with Marilyn to discuss each other's reasoning, both women rank *believing R* as less OFF than *withholding judgment about R*. (R, recall, is the proposition that *the prize is behind the rightmost door*.) After consulting with each other, neither acquires any higher-order evidence to support revising this ranking. In accordance with *Unanimity*, any pure HOE view must thereby entail that the rational response to the women sharing their reasoning with each other is for both women to retain their original ranking of *believing R* as less OFF than *withholding judgment about R*.

5. An Arrovian-Style Impossibility Theorem

Unanimity and *IIA* jointly entail that *Fallibility* is false. I shall call this result *Arrow's Epistemological Theorem*. Since conciliatory pure HOE views entail both *Unanimity* and *Fallibility*, and since the motivation for *IIA* is that updates to response function outputs after full disclosure should be proportionate to changes epistemic peers make to their doxastic profiles after full disclosure (if, indeed, such outputs should be updated at all), this theorem amounts to the claim that conciliatory pure HOE views demand disproportionate updates of response function

output when peers change their doxastic profiles after full disclosure despite acquiring no new evidence (and losing no available evidence).

The proof of Arrow's Epistemological Theorem, following Geanakoplos, involves three steps.¹⁵ The first shows that, for *any* doxastic attitude Y, if, after full disclosure, everyone in a peer group ranks Y as either not less OFF than anything else or less OFF than everything else, then the response function must rank Y as either not less OFF as anything else or less OFF than everything else. The second shows that, for a *particular* doxastic attitude Y, there is someone in the peer group who is *infallible* with respect to all pairwise rankings not involving Y. The third step shows that this same person must be infallible with respect to all pairwise rankings, regardless of whether they involve Y.

Step 1. Consider a situation in which, after everyone has disclosed to one another the reasons for their attitudes toward some arbitrary proposition, all epistemic peers have doxastic profiles that rank some arbitrary doxastic attitude Y toward that proposition as *either* not less OFF than any other attitude or less OFF than all other attitudes: after full disclosure, everyone's profile has either Y ON and other attitudes OFF, or Y the most OFF of all attitudes. (This situation might be one in which half of the peers rank Y as not less OFF than any other attitude, while the other half rank Y as less OFF than all other attitudes.) *IIA* and *Unanimity* entail

Extremal Lemma: For any doxastic attitude Y toward a proposition and any peer set of doxastic profiles for that proposition, whenever every peer ranks Y as either not less OFF than any other attitudes or less OFF than all other attitudes after full disclosure, the output of any response function must either rank Y as not less OFF than any other attitude or else rank Y as less OFF than all other attitudes.

Suppose, for *reductio*, that the response function does not rank Y in either of these ways. Then there are attitudes X,Z such that the response function yields, as output, that X should be less OFF than Y and Y should be less OFF than Z. Now suppose that, for whatever arbitrary reason and despite no change in available evidence, every peer's doxastic profile changes so that each person ranks Z as less OFF than X while not changing their pairwise rankings involving Y. Then *IIA* entails that the response function continues to yield, as output, that X should be less OFF than Y and Y should be less OFF than Z. Transitivity of the *is less OFF than* relation entails that this function yields that X should be less OFF than Z. However, *Unanimity* entails that the function yields that Z should be less OFF than X. Discharging the contradiction and completing the *reductio* establishes the lemma.

Step 2. Next, consider a particular doxastic attitude Y toward a proposition and a situation in which all peers have doxastic profiles that rank Y as more OFF

¹⁵ Geanakoplos, "Three Brief Proofs," 212-213.

than all other attitudes after full disclosure (otherwise the rankings in the peer profiles are arbitrary). Call this Situation 1. Imagine that, for whatever arbitrary reason and despite no change in available evidence, each of N peers successively changes her profile so that Y goes from being ranked as *more* OFF than all other attitudes to being ranked as *less* OFF than all other attitudes. Let Situation N be the situation, after full disclosure, in which all peers have doxastic profiles that rank Y as less OFF than all other attitudes. In Situation 1, *Unanimity* entails that the output of the response function should rank Y as more OFF than all other attitudes. The Extremal Lemma entails that, for every situation between Situation 1 and Situation N , the response function should either rank Y as more OFF than all other attitudes or else rank Y as less OFF than all other attitudes. In Situation N , *Unanimity* entails that the output of the response function should rank Y as less OFF than all other attitudes. Clearly, there must exist a peer, n^* , whose profile change causes a change in the output of the response function.

Let Situation A be one in which this n^* has a doxastic profile that ranks Y as more OFF than all other attitudes, and let Situation B be like Situation A except that n^* has changed to have a profile that ranks Y as less OFF than all other attitudes. Then the output of the response function in Situation A should rank Y as more OFF than all other attitudes; and in Situation B , it should rank Y as less OFF than all other attitudes. Consider two arbitrary doxastic attitudes X, Z , each distinct from Y , and construct an arbitrary Situation C from Situation B that satisfies the following conditions:

- the profiles for peers 1 through n^*-1 rank Y as less OFF than any other attitude,
- the profiles for peers n^*+1 through N rank Y as more OFF than any other attitude, and
- the profile for n^* ranks X as less OFF than Y and Y as less OFF than Z .

IIA entails that output of the response function regarding the relative ranking of X and Y for Situation C should be the same as it is for Situation A . Given the relation between Situation A and Situation B , Situation A and Situation C have the same pairwise rankings of X and Y for all peer profiles. Since, in Situation A , the output of the response function is that X should be less OFF than Y , *IIA* entails that this is the output of the response function in Situation C as well. Similarly, given the relation between Situation B and Situation C , those situations have same pairwise rankings of Y and Z for all peer profiles. Since, in Situation B , the output of the response function is that Y should be less OFF than Z , *IIA* entails that this is the output of the response function in Situation C as well. Transitivity of the *is less OFF*

than relation thereby entails that, in Situation C, the output of the response function should rank X as less OFF than Z.

A similar argument shows that if, in Situation C, the doxastic profile for n^* were to rank Z as less OFF than Y and Y as less OFF than X, the output of the response function in Situation C would rank Z as less OFF than X. Hence, for a *particular* doxastic attitude Y, there is an n^* in the peer group who is infallible with respect to all pairwise rankings not involving Y, in the sense that this person determines the response function's output for how those alternatives should be ranked. A similar argument, considering a *different* particular doxastic attitude Z, shows that there is also a person, n^{**} , in the peer group who is infallible with respect to all pairwise rankings not involving Z.

Step 3. Suppose, for *reductio*, that n^* is not the same person as n^{**} . Then n^* cannot affect the response function's output regarding the relative ranking of alternatives X and Y, because n^{**} determines that output. Yet clearly sometimes n^* does affect this output, as with Situations A and B. Hence, $n^*=n^{**}$. Similar arguments show that n^* determines the response function's output for *all* rankings, and this amounts to *Fallibility* being false.

Therefore, if *Unanimity* and *Fallibility* are true, *IIA* is false. This is Arrow's Epistemological Theorem, and it places a burden on advocates of conciliatory pure HOE views.

If they accept the theorem, their burden is to show that, when updates to response function outputs after full disclosure are *not* proportionate to changes epistemic peers make to their doxastic profiles after full disclosure, the updated outputs continue to capture rational responses to evidential situations among epistemic peers. If they reject the theorem, their burden is to show that some background presupposition for the theorem fails.

I maintain that updates to response function outputs are rational only if they are proportionate, so that any view that denies *IIA* is false. So far as I know, the extant literature on peer disagreement does not provide an argument to the contrary. Accordingly, if Arrow's Epistemological Theorem is sound, it shows that no pure HOE view can be conciliatory. For pure HOE views endorse both *Unanimity*, conciliatory views endorse *Fallibility*, and the theorem shows that *Unanimity* and *Fallibility* jointly entail that *IIA* is false.

6. Prospects for Avoiding Arrow's Epistemological Theorem

If no pure HOE view can be conciliatory, one of the most popular views about peer disagreement, the equal weight view, must be mistaken. Since many epistemologists have strong intuitions that something like the equal weight view must be true, it is

worth considering some options for rejecting Arrow's Epistemological Theorem. The theorem, after all, requires several background presuppositions, and if one of these were to be false, the theorem would not be sound.¹⁶ I shall consider the prospects for denying four such presuppositions, arguing that each prospect is unpalatable for those who accept views about peer disagreement that are both conciliatory and pure HOE.

An advocate of a conciliatory pure HOE view might object that, even if such a view may be understood as maintaining that there is a function that takes as input information about higher-order evidence about disagreeing peers' doxastic attitudes toward a disputed proposition after the peers disclose to each other the reasons for their conflicting attitudes and yields as output a verdict about the rational response to that disagreement, the output of this function is not a ranking of doxastic attitudes in terms of the *is less OFF than* relation. Instead, the objection might go, the output of a response function is merely information about which particular attitude(s) peers ought to adopt toward a proposition after full disclosure. This is output about which attitude(s) should be the least OFF one(s). However, even if this is correct, analogues of Arrow's Epistemological Theorem hold under reasonable conditions.¹⁷ So this option does not seem promising.

Rather than focusing on outputs of response functions, an advocate might focus on inputs, objecting that response functions need take as input *only* information about which peer doxastic attitudes happen to be ON after the peers disclose to each other the reasons for their conflicting attitudes. After all, the argument against this understanding of response function input relies upon a special kind of case, namely, one in which disagreeing peers adopt conflicting doxastic attitudes toward a proposition after full disclosure and there is no "middle-ground" doxastic attitude for them toward which they can move. That conciliatory pure HOE views fail to handle this kind of case does not show that they do not handle *any* kind of peer disagreement. Hence, this objection goes, even if Arrow's Epistemological Theorem shows that conciliatory pure HOE views are false when applied to a special kind of case, it does not show that such views are false more generally.

While this objection is cogent, it rescues conciliatory pure HOE views about disagreement from refutation at the cost of making them unattractively ad hoc. If advocates of conciliatory pure HOE views opt to restrict the range of cases to which such views apply, then, in the special cases, either some peer need not change her

¹⁶ I am indebted to the discussion of some of these presuppositions in Samir Okasha, "Theory Choice and Social Choice: Kuhn versus Arrow," *Mind* 120 (2011): 83-115.

¹⁷ See Austen-Smith and Banks, *Positive Political Theory I*, 49-52.

doxastic attitude toward the others or else first-order evidence helps to determine how the peers should change their attitudes. But there does not seem to be a principled reason for allowing that a steadfast response is rational when there is no "middle-ground" attitude and yet denying that a steadfast response is rational when there is, because facts about how many possible doxastic attitudes happen to be available between two peers' conflicting attitudes are not facts about higher-order evidence (thereby violating the spirit, if not the letter, of pure HOE views), and because such facts do not seem to be relevant to the rationality of a response to peer disagreement. Moreover, maintaining that there is a default doxastic attitude, such as withholding assent, removes the appearance of adhocery by virtue of not being a conciliatory view. For if, say, the proper rational response to disagreement after full disclosure between a theist and an agnostic is for both to withhold assent about whether God exists, the agnostic's doxastic attitude remains unchanged.

Perhaps, however, advocates of conciliatory pure HOE views can avoid the charge of adhocery by denying that the special cases pose any problem at all. The argument that they do depends upon the claim that possible doxastic attitudes for actual people are not dense. But, one might object, an advocate of a conciliatory pure HOE view need not be moved by this contingent fact, because the claim to the contrary may be understood as an idealization, and idealized theories do not merit any special concern. For example, even though the equation of motion for the simple pendulum is idealized by virtue of treating the pendulum bob as a point-mass particle and the pendulum string as perfectly rigid (among other things), the equation remains useful and legitimate to use for certain situations in which these idealizing conditions do not obtain.

There is something correct about this objection. Idealized theories often are not particularly worrisome. Nonetheless, the objection is flawed. The idealizations that do not cause concern are *controllable*: there is some way to take into account the distorting effects of the idealization.¹⁸ This accounting might involve removing the idealization, showing that its effect on the theory is negligible, and so on.¹⁹ However, the density idealization is not controllable, because response functions for density-idealized conciliatory pure HOE views produce outputs that their counterpart non-density-idealized response functions deem to be impossible. There is no way to remove the density idealization, or to estimate the idealization's effect

¹⁸ See Lawrence Sklar, *Theory and Truth: Philosophical Critique within Foundational Science* (New York: Oxford University Press, 2000), 44, 61-70.

¹⁹ For example, see Ronald Laymon, "Idealization, Explanation, and Confirmation," *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume One (1980): 336-350.

because, without the idealization, response function outputs for conciliatory pure HOE views are either incorrect or inapplicable to the actual world. In this respect, the density-idealization is akin to the idealization of systems as having infinitely many particles in statistical mechanical accounts of phase transitions.²⁰ When idealizations are uncontrollable, it is not clear that theories which rely upon them have any applicability to the real world. If they do not, such theories might be true of some idealized situations, but they are false of real ones.

A fourth way for an advocate of a conciliatory pure HOE view to avoid Arrow's Epistemological Theorem is to maintain that *Depth* is false. If one doxastic attitude can be less OFF than another only when the former is ON and the latter is OFF, the proof of Arrow's Epistemological Theorem fails. However, *Depth* is an extremely weak claim. Its truth is compatible with all actual people's rankings of doxastic attitudes being such that one attitude is less OFF than another only when the former is ON and the latter is OFF. Even if thinking of doxastic attitudes as being ON or OFF and ranking doxastic attitudes in terms of the *is less OFF than relation* is new, this novelty alone does not support the strong modal claim that *Depth* is false, especially when the Ranking case provides at least some evidence to the contrary.

The responses to the preceding objections suggest that conciliatory pure HOE views about peer disagreement are false if they cannot avoid Arrow's Epistemological Theorem (by virtue of violating *IIA*) and that they can avoid Arrow's Epistemological Theorem only by virtue of being unacceptably ad hoc. There are other views about peer disagreement that can accept the theorem without being ad hoc and without violating *IIA*. But these are unpalatable for views that are both conciliatory and pure HOE, because they involve adopting views that are either steadfast or mixed. For example, consider

Extreme. Two rationally competent peers mistake the import of a shared body of evidence regarding hypothesis H. In response to the evidence, one peer gives credence 0.7 to H and the other gives it 0.9. However, the evidence *in fact* supports only the credence 0.3 for H.²¹

Kelly takes this kind of case, in which disagreeing peers radically misevaluate the import of their evidence, to show that pure HOE views are incorrect. Suppose he is right. But suppose that these kinds of cases support a view according to which,

²⁰ See Craig Callender, "Taking Thermodynamics Too Seriously," *Studies in History and Philosophy of Science Part B* 32 (2001): 539-553; Chuang Liu, "Infinite Systems in SM Explanations: Thermodynamics Limit, Renormalization (Semi-) Groups, and Irreversibility," *Philosophy of Science* 68 (2001): S325-S344.

²¹ This adopts Case 5 in Kelly, "Peer Disagreement and Higher-Order Evidence," 125-126.

when all peers have the same attitude toward a proposition, the rational response to the evidence for that proposition is a function of first-order evidence *only*. This is a mixed view, and it entails that *Unanimity* is false. For even if everyone were to mistake the import of the evidence for a hypothesis and adopt the same incorrect credence toward that hypothesis after full disclosure, the rational response to the evidence would not be to adopt that particular mistaken credence.

7. Concluding Remarks

Whether a person has a particular doxastic attitude toward a proposition is not an all-or-nothing affair. For there are situations in which one doxastic attitude for a person toward a proposition can be less OFF than another attitude of that person toward the same proposition (see Section 2). The extant literature on the epistemology of peer disagreement overlooks this kind of depth in people's doxastic attitudes. But acknowledging this depth allows conciliatory pure HOE views of peer disagreement to avoid certain difficulties, by virtue of denying that the information about doxastic profiles of epistemic peers taken as input by response functions for such views is merely information about which doxastic attitudes happen to be ON for those peers (see Section 3).

An adequacy condition for response functions that take as input more information than information concerning which doxastic attitudes happen to be ON for epistemic peers is *IIA* (Independence of Irrelevant Alternatives): for any pair of distinct doxastic attitudes toward a proposition, if some epistemic peers change their doxastic profiles toward that proposition after full disclosure, without changing the relative ranking of those doxastic attitudes, the output of the response function does not change the relative ranking of those attitudes either. This condition ensures that updates to response function outputs do not change relative rankings of doxastic attitudes without reason (see Section 3). Conciliatory pure HOE views impose additional constraints on response functions (see Section 4). Yet, according to Arrow's Epistemological Theorem, these constraints are jointly incompatible with *IIA* (see Section 5). Accordingly, given *IIA*, if Arrow's Epistemological Theorem is sound, no pure HOE view of peer disagreement can be conciliatory and, in particular, the popular equal weight view is mistaken.

While there are ways to avoid Arrow's Epistemological Theorem, none of them should be appealing to advocates of conciliatory pure HOE views (see Section 6). Restrictions on the output of response functions succumb to analogues of the theorem. Restricting the inputs of response functions makes conciliatory pure HOE views either ad hoc or inapplicable to real cases, thereby preventing them from providing a general answer to the question of what disagreeing peers ought to do.

An Arrovian Impossibility Theorem for the Epistemology of Disagreement

Finally, rejecting certain constraints on response functions themselves involves adopting views about peer disagreement that are either steadfast or mixed.²²

²² I thank George Schumm and participants at the 2011 Omaha Epistemology Workshop for helpful comments on earlier versions of this paper.