The Wisdom of the Crowd: The Evidential Role of Convergence and Consensus

1. Purpose and aims

An intense debate in epistemology concerns how one should adjust one's beliefs upon learning about the assessments of those beliefs by other people. Much of this debate concerns disagreement and when, if ever, the fact that others reject one's beliefs is a reason to reduce one's confidence in them. However, there are important questions also about when others share one's views. For example, if dissent is sometimes a reason to be more skeptical, might agreement correspondingly be a reason to increase one's confidence?

As this question suggests, disagreement and agreement may be seen as two sides of the same coin. Disagreement provides a potential skeptical challenge, while agreement provides potential assurance. Yet the attention that philosophers have devoted to these phenomena is quite unevenly distributed: while there is a rich and growing literature about the epistemological significance of disagreement, the significance of agreement has been much less discussed.

What explains the asymmetry? Perhaps the assumption that disagreement has a kind of primacy, in the sense that agreement is significant only because it can play a negative, deflective role, by signifying that one type of skeptical challenge fails to apply. Alternatively, the thought could be that the theories that have emerged in the discussion about disagreement could easily be extended to cover agreement as well, and that as a result, there is no need for a separate treatment. Both these ideas are disputable, however, and it is not clear that they justify the lack of independent explorations of the significance of agreement. Indeed, before such investigations are carried out, the ideas are even hard to properly evaluate. The limited attention paid to agreement within epistemology can accordingly be seen as a lacuna, and the purpose of the project is to help fill that lacuna, by making agreement its direct object of study.

More specifically, we wish to explore whether agreement, besides playing the deflective role just indicated, can also play a more offensive one, by providing *positive* justification for belief in the shared claims. The central question can be stated: Under what circumstances, if any, does agreement among thinkers provide a reason to trust the shared beliefs? By addressing this question, we will contribute to several central debates within epistemology. However, the project also has a wider relevance. There is a growing concern that the existing scientific consensus about claims that motivate urgent policy decisions is not treated with the respect that it deserves. In particular, an increasing number of individuals routinely dismiss the consensus by groundlessly attributing it to conspiracies or sociological mechanisms that allegedly do not track the truth, or by simply insisting that "truth is not a matter of voting". Such objections form a part of the dialectical arsenal of climate deniers and proponents of the anti-vaxxer movement, for example, and to assess and respond to them, one needs a better understanding of how arguments from agreement are supposed to work. The aim of the project is to contribute to such an understanding.

2. State-of-the-art

The project belongs to the growing field known as "social epistemology". Many debates in this field are relevant to the questions that we wish to address, including those that concern peer disagreement, testimonial knowledge, Condorcet's jury theorem, and expertise. However, the project will also draw on other bodies of literature, such as those that study formal accounts of convergence in the context of Bayesian confirmation theory and belief revision, and those that focus on "higher-order evidence" and irrelevant influences on beliefs.

The peer disagreement debate concerns how we should respond when finding that our beliefs are opposed by someone we have reason to regard as a "peer", where a peer, roughly, is someone who has access to the same evidence, is equally good at reasoning, and is generally just as well-equipped as we are relative to the aim of determining the truth about the relevant topic (see e.g. Christensen 2009 and Elga 2007). A number of views have emerged, including both so-called "conciliationist" ones, which imply that we, when facing such dissent, should reduce our confidence in the target belief (Christensen 2007, Elga 2007,

Feldman 2006), and those that instead permit "steadfastness" (Enoch 2010a, Kelly 2011; the debate is interestingly approached from a Bayesian perspective in Mulligan forthcoming). However, whether the results from this debate can straightforwardly be extended to cases of agreement is unclear. For example, while parties in the debate tend to assume that disagreement with epistemic "inferiors" (i.e., people who are less well equipped than us) does not command reducing one's confidence in the truth of the target belief (see, e.g., Kornblith 2010), there is reason to think that agreement with inferiors may potentially provide support.

One possible argument to that effect invokes Condorcet's jury theorem (hereafter "the jury theorem"), which applies to cases where a group of people ponder a question which can be correctly answered with a yes or a no. The original theorem states that the probability that the majority's verdict is true approaches certainty as the size of the group increases, provided that their verdicts are relevantly independent and that each has a chance (the same for all individuals) of being correct that is greater than .5 (Condorcet 1785; see also Grofman et al 1983). Since a group of people could satisfy both conditions and still be our inferiors, the jury theorem suggests that agreement with inferiors may also have considerable weight.

The jury theorem's relevance to actual cases of agreement is disputed, however. For example, in actual cases, individuals often have more than two options to consider, and interact in ways that may be hard to reconcile with the independence condition. The homogeneity assumption that all have the *same* chance of being correct is also obviously problematic. At the same time, however, a number of generalizations of the jury theorem have been developed which apply also to non-binary assessments (List and Goodin 2001) or derive similar conclusions about the majority's reliability from assumptions that allow for certain forms of dependence (Boland 1989, Kaniovski 2010 and Pivato 2013) or for variance regarding the individuals' levels of competence (Berend and Sapir 2005, Dietrich 2008).

One question regarding the jury theorem's applicability concerns the reasons we may have for believing that a person is sufficiently competent. What sorts of facts could constitute such reasons? This and related questions are addressed in the discussion about testimony and expertise (for an overview, see the essays in Lackey and Sosa 2006). Several considerations are clearly relevant, including the testifier's training and past track record. A less well explored factor is addressed by Goldman (2001) in a discussion about when we should treat somebody as an expert. According to Goldman, this depends on the extent to which her views are shared by other (putative) experts in the field. Goldman warns, however, that agreement need not always indicate reliability. For instance, he notes (in a way that is reminiscent of the jury theorem) that the views of the agreeing parties might not be sufficiently independent, as the parties might have been "trained by one and the same 'guru', who was a very persuasive and compelling figure" (see also Barnett 2019). What Goldman here highlights is that the significance of agreement depends in part on how it is to be explained, and that some explanations undermine the thought that the agreement is an indication of truth.

What would an explanation that instead *vindicates* such a thought look like? One approach to this question appeals to formal accounts of scientific convergence, such as the Lehrer-Wagner and the Hegselmann-Krause models (Lehrer and Wagner 1981, Riegler and Douven 2009). Another relevant debate focuses on irrelevant influences and debunking arguments (White 2010, Vavova 2018). That debate partly concerns the question of how we should respond when we learn that some of our beliefs are best explained by accounts that do not assume their truth (Joyce 2006, Street 2007, Tersman 2017). While some writers take that fact to undermine the justification of the pertinent beliefs, others argue that such an explanation may still attribute the beliefs to factors that are indirectly related to their truth in a way that leaves their justification intact (Enoch 2010b, Clarke-Doane forthcoming). How to apply those views to explanations of agreement is a question that deserves further exploration.

Another relevant body of literature addresses the significance of disagreement and irrelevant influences on belief under the assumption that they provide "higher-order evidence", or "HOE". While it is controversial how the concept of HOE is best defined, the idea is that it differs from "first-order" evidence in that it is relevant to a belief's justification without bearing directly on its truth. (For two alternative understandings of HOE, see Lasonen-Aarnio 2014 and Kelly 2005.) Suppose, for example, that we learn that our belief that Minsk is the capital

of Belarus was formed under the influence of hypnosis. While this information (which is an instance of HOE) hardly indicates that Minsk is *not* the capital of Belarus, it may still seem to mandate reduced confidence in the truth of the belief. The debate about HOE concerns whether (and, if so, why) that is in fact the case. A number of views have been suggested, including ones on which HOE may undermine the justification of a belief by silencing the pertinent first-order evidence in its favor (Christensen 2010, Horowitz 2014), ones on which HOE instead has to be weighed against the subject's first-order evidence (Kelly 2011), and the so-called "level-splitting" approach, on which HOE may impact the justification of second-order beliefs about whether one's first-order beliefs are justified, but never impacts the justification of the first-order beliefs themselves (Lasonen-Aarnio 2014, Worsnip 2018).

3. Project description

Recall the main question of the project: Under what circumstances, if any, does agreement among thinkers provide a (positive) reason to trust the shared beliefs? As far as we can see, there is no reason to rule out that there may be several plausible answers to that questions; answers that are distinct but not inconsistent with each other. There may also be distinct but mutually consistent answers to the question of *why* agreement is thus relevant. In part, this is because different explications of the concept of agreement may be relevant in different contexts. We shall here adopt a "coarse-grained" one, on which what matters is simply whether the parties have the same beliefs, but one can also imagine more complex definitions (such as ones that take degrees of belief into account; see further Rowbottom 2018).

We shall approach our main question by examining two distinct ways of arguing that agreement can play such a positive role. According to arguments of the first kind, agreement provides justification in some cases (partly) because the agreeing parties have reached their shared beliefs independently of each other. The availability of this reasoning is illustrated by Condorcet's jury theorem, which suggests that agreement among independent thinkers can sometimes be direct evidence for the truth of the shared beliefs. As we explain below, however, there may also be versions of the strategy that rely on other premises.

According to arguments of the second kind, agreement can provide justification in virtue of being the outcome of a certain sort of collective deliberation process. The source of inspiration is a certain picture of how competent scientific enquiry is supposed to proceed. Research ideally takes place in an environment in which dissent is tolerated (and indeed encouraged) and researchers may therefore initially have conflicting beliefs. However, by exchanging arguments, exposing their beliefs to criticism, and by jointly exploring novel ideas, they sometimes manage to resolve their disagreements. The idea is that if a theory prevails in such an environment, and if the initially disagreeing parties thus converge upon it, then that might also be a powerful reason to accept the theory. This line of thought is prominently associated with, among others, John Stuart Mill (1859, 41 & 95).

One aim of the project is to examine and evaluate the chances of success of these strategies. Another is to explore the relations between them. The type of interactions central to the second strategy are often taken to compromise the independence of the agreeing beliefs in the sense relevant to the first strategy. This may seem to indicate that the strategies are mutually exclusive in the sense that they apply (if at all) only to different cases. However, it is disputed how the relevant concept(s) of independence should be understood, and one of the ideas that we will investigate is that there may be interpretations that allow for successful combinations of the strategies.

An improved understanding of the two strategies is also likely to shed light on another central question in the project, which concerns the relation between arguments from agreement and arguments from disagreement. We noted earlier that agreement has received significantly less attention among philosophers. One possible reason is that many assume that an adequate theory about the significance of disagreement will easily extend to agreement (see, e.g., Lutz 2019). Although this symmetry idea may initially seem appealing, however, it has not been subject to sufficient scrutiny. A further aim of the project is to provide such an evaluation, by exploring symmetries and asymmetries between agreement and disagreement.

To achieve our aims, the project will be divided into three parts. The more specific topics that shall be addressed within each part, as well as the hypotheses that will guide our investigations, are described below.

3.1 Agreement and Disagreement – Symmetries and Asymmetries
This part of the project will assess how arguments from agreement and arguments from disagreement interact. The aim is to evaluate what we shall call "the symmetry thesis", according to which a correct theory about significance of disagreement will straightforwardly generate answers to the parallel questions about the significance of agreement as well. We shall test this thesis by considering some potentially important contrasts in how agreement and disagreement might serve as evidence.

The upshot of arguments based on the jury theorem is that agreement sometimes can be direct, "first-order" evidence for a belief. The same holds, arguably, for disagreement (see, e.g., Kelly 2011). In addition, however, disagreement is often taken to play other roles as well. On one suggestion, for example, finding that one's beliefs are opposed by others can provide what Pollock (1986) has termed "undercutting defeat", where this means that the primary object of the challenge is not the target belief itself but some background belief that links the belief to the evidence on which it is based (see Risberg and Tersman 2020). On a related suggestion, disagreement can provide "higher-order evidence" (or "HOE") against a belief, which is supposed to mean that it undermines the belief without bearing directly on its truth but also without targeting some background linking belief (see, e.g., Christensen 2010). Can agreement also play corresponding additional roles? If so, is there nevertheless room for differences in the ways the two phenomena can be relevant to justification (beyond the obvious one that disagreement presumably affects justification negatively while agreement does so positively)? Those are two of the questions that we shall address.

The second question raises the more general issue of how positive and negative HOE relate to each other, which is an underexplored issue in the literature. The debate has focused almost exclusively on negative HOE, perhaps because a uniform treatment of positive and negative HOE is assumed. Such a uniformity thesis about positive and negative HOE is not obviously true, however, and one of our aims in this part of the project is to challenge it by examining the nature and significance of positive HOE more closely.

While the exact definition of negative HOE is a matter of controversy, the concept has been used to cover considerations of at least two (quite different) types:

(i) evidence that our faculties are not functioning properly (Lasonen-Aarnio 2014);
(ii) evidence that we have evaluated our first-order evidence incorrectly (Kelly 2005).

A uniformity thesis about positive and negative HOE should thus lead us to expect that there will also be at least two different types of positive HOE:

(i*) evidence that your faculties are functioning properly; and
(ii*) evidence that you have correctly evaluated your first-order evidence.

However, while the view that both types of negative HOE calls for reduced confidence is a serious contender, the corresponding view about positive HOE—that both types of positive HOE call for *increased* confidence—faces serious problems. In particular, while it is plausible that positive HOE of type (i*) might call for such a reaction, HOE of type (ii*) does not always seem to do so. Suppose, for example, that after studying some complicated weather data, you reach a certain degree of confidence in the claim that next year will be unusually rainy. If a weather expert lets you know that your evaluation of the data was correct, in the sense that the data in fact justifies that degree of confidence, it would clearly be unreasonable of you to become *more* confident that next year will be unusually rainy. The fact that the expert agrees with your assessment of the data may give you a reason to think that your assessment was correct, of course, and it is thus naturally seen as an instance of positive HOE of type (ii*). But

it does not seem to give you a reason to change your confidence about the first-order question about next year's weather. (If anything, it gives you a reason *not* to change your confidence.)

What this suggests, we think, is that a view on which positive and negative HOE affects justification in different ways at least deserves to be taken more seriously than what has so far been noted. Given the idea that agreement provides positive HOE while disagreement provides negative HOE, then, such a view would also entail that agreement and disagreement are relevant to justification in different ways. And this upshot would, in turn, cast doubt upon the symmetry thesis about agreement and disagreement.

That said, a more detailed assessment of the symmetry thesis requires considering several further issues. One set concerns the different ways in which agreement can affect the justification of a belief. When, if ever, should it be seen as providing positive HOE, rather than first-order evidence? Moreover, if agreement sometimes does provide HOE, under what conditions is the HOE best seen as being of type (i*) and/or (ii*), respectively? These questions interact in different ways with those addressed in other parts of the project. For example, the robustness approach considered earlier raises the question of whether the robustness of a belief should be taken to provide first-order or higher-order evidence in its favor. Other questions concern how a theory of HOE should be designed to accommodate the different kinds of support that agreement, and positive HOE more generally, may provide. For example, might there be forms of HOE other than those considered above, and if so, what responses do they call for? And how do those forms relate to other considerations that are also thought to impact justification indirectly, such as undercutting defeaters?

3.2 Agreement Among Independent Thinkers
As noted above, one way to argue that agreement among independent thinkers may be evidence for the truth of the shared beliefs proceeds via Condorcet's jury theorem. A problem with the jury theorem, however, is that its applicability to actual cases is often dubious. Some of the worries are due to the competence condition, which requires, for example, that all individuals have the same (greater than .5) probability of being correct. Moreover, although there are generalizations of the theorem that involve more realistic assumptions (see, e.g., Berend and Paroush 1998), there are also worries that emanate from the independence condition. For example, it has been argued that this condition interacts with the competence condition in such a way that, for many actual cases of agreement, there are no interpretations of the conditions such that both of them can justifiably be held to be satisfied (Dietrich 2008).

These difficulties raise the question of whether there are promising ways to argue that independence matters which do not proceed via the jury theorem. To develop and examine such alternatives is the aim of this part of the project. One of the ideas that we intend to probe can be indicated as follows: Reconsider Goldman's guru case (see section 2). Why should we be unimpressed by agreement that is due to the fact that one of the individuals is treated by the others as a guru? A simple idea is that if the disciples just mechanically accept the guru's teachings, then their affirmation of her beliefs does not add any credibility to those beliefs beyond that which is afforded by the affirmation of the guru herself. Whether the guru's own affirmation affords such credibility depends on the concerns we may have about her reliability. What is clear, however, is that whatever those concerns happen to be, they are directly inherited by the beliefs of her followers. So, by noting that the guru's beliefs are shared by her devout followers, we do not acquire any additional reason to trust them.

In searching for an epistemically relevant concept of independence, one might therefore propose that the beliefs of agreeing individuals are independent to the extent that they are *not* vulnerable to the same reliability concerns. The individuals may have reached their beliefs on the basis of different sets of data, for example, or by using different methods. Why is agreement among thinkers who are independent in this sense supposed to matter? One answer invokes the concept of "inferential robustnesss", which has been developed in a different context (Woodward 2006; see also Weisberg 2006, Wimsatt 2011, and Schupbach 2018) and signifies a property which a conclusion has insofar as it can be derived from a given body of evidence under a variety of different background assumptions. The inferential robustness of a conclusion arguably decreases the "cognitive risks" involved in accepting it,

since its viability is not as sensitive to objections to the particular set of background beliefs from which one may have inferred it. The risks seem even lesser when the conclusion is also robust with respect to different sets of data or to different belief-forming methods. The idea that we shall pursue is inspired by these thoughts, by taking agreement among independent thinkers to matter because it suggests that the shared beliefs have such robustness. In what follows, the term "methodological robustness" is used to cover all the indicated forms.

To evaluate this idea, several issues need to be addressed. For example, the relevant concept of independence must be clarified. What does it mean, more specifically, to say that the beliefs of different individuals are subject to the "same" reliability concerns? Answering this question involves, among other things, considering when different individuals have used the same or different belief-forming methods. The issue of how such methods should be individuated has been given some attention in the literature (Becker 2012, Zhao 2018), but it is usually approached in an individualistic way, as concerning when one and the same thinker has used the same method at different occasions. One task in this part of the project is to explore the implications of those discussions in the present context.

Another set of questions concerns *how* exactly the methodological robustness of a belief is supposed to be relevant to its justification. On the basis of the jury theorem, it is possible to argue that agreement is in some cases evidence of the truth of the shared beliefs. Can the methodological robustness of a belief also provide such evidence? Or is it at best relevant in a more indirect way, by influencing the justification of the belief without bearing directly on its truth (in the way higher-order evidence is supposed to do)?

Finally, a third set of issues concerns the relation between the concept of independence associated with the robustness approach and that which is invoked by the jury theorem. The latter concept is often taken to rule out many factors that are typically present in actual situations, such as shared information, opinion leaders, and group deliberation (Dietrich and List 2004, Owen 1986). Does this hold also for the alternative concept of independence outlined above as well, or can shared information (and so on) more easily be reconciled with the supposition that the individuals are not vulnerable to the same reliability concerns?

3.3 Convergence and Collective Deliberation
The second part will be devoted to arguments that focus on agreement that results from a process of the kind that is often supposed to be instantiated in the sciences. One of its crucial features is the interactive aspect. The beliefs which the subjects have converged upon have been exposed to testing in the form of comparisons with alternative views and of discussions between trained individuals. Since selection is a crucial element, another important feature is that the process takes place in a context where there are a plurality of competing views. That is why those who favor the collective deliberation-based strategy, such as Mill, stress the importance of fostering a culture that encourages dissent. The aim is to create a "marketplace of ideas" (e.g., Zamora Bonilla 2012). Note that although the sciences provide the prime example of this process, the belief in its truth-conduciveness can be traced also in the design of the so-called "Delphi method" (Burgman 2005) and similar methods used in forecasting.

How may the (alleged) truth-conduciveness of the process just sketched be explained and how could it be promoted? These are complex questions which can be approached by consulting both empirical studies (see, e.g., Clemen and Winkler 1999 and the references therein) and the formal accounts of convergence mentioned above (section 2). According to one idea, the interactive aspect is significant because it enables the "pooling" of cognitive resources. Some people are good at coming up with novel theories, others are good at criticizing them, and so on. A central issue in this part of the project is whether that aspect rules out the type of independence that is stressed by the strategy explored in the second part. In discussions of the jury theorem, several theorists have taken for granted that interaction of the indicated type violates the relevant independence condition (see, e.g., Rawls 1971: 538, Grofman et al 1983, and Hedden 2017). On closer inspection, however, things are more complex (see. e.g., Waldron's contribution to Estlund et al 1989: 1327). A response of particular interest to us is offered by Estlund, who stresses the distinction between

communication and deference and argues that it is the latter, not the former, that independence should be taken to exclude:

> [W]here there is no communication, there certainly can be no deference nor any of its ill effects. However, if deference can be avoided, communication would seem to have advantages from a Condorcetian perspective. Increasing the information of a nondeferential voter would tend to increase the voter's competence, and this can only increase the chances of the group's ascertaining the general will [or the truth] (in Estlund et al. 1989: 1320).

The weight that Estlund assigns to deference is congenial with the diagnosis of Goldman's guru-case considered above. On that diagnosis, if the guru's followers unthinkingly accept the guru's beliefs then their affirmation of them lacks significance because whatever worries we may have about the guru's reliability apply to them as well. However, this seems to hold only if the deference is complete, or "blind", in the sense that it leaves no room for ways in which the deferring individuals can make a distinct contribution to the assessment of the target beliefs. Deference which is merely partial in some relevant dimension might be less threatening, as Estlund acknowledges in a later paper (1994). Which dimensions are thus relevant and how is the concept of (blind) deference more specifically to be defined? These are two of the questions that we shall address.

The robustness approach may for several reasons be better placed than strategies that invoke the jury theorem to accommodate the epistemic value of the deliberative process instantiated by the sciences. For example, it provides a seemingly straightforward explanation of the significance assigned to the fact that the agreement has emerged against a background of disagreement. The initial disagreement suggests differences in background beliefs or methodological commitments such that if agreement emerges in spite of these differences, then this suggests that the relevant conclusions have at least some degree of methodological robustness. Other aspects of the process also seem easier to handle on the robustness approach, such as the sharing of evidence that the interactive aspect entails. When evidence is shared it becomes a common influence of the participating individuals' beliefs, and such influences have also been taken to exclude independence of the form relevant to the jury theorem (see section 3.2). However, it is less clear that they violate the independence requirement associated with the robustness approach, as they do not exclude that the individuals use different background beliefs to derive their assessments of the shared beliefs from that evidence. These potential advantages of the robustness approach shall be considered in more detail in this part of the project.

References
Barnett, Zach. 2019. "How do numbers count?", *Philosophical Studies* 176: 297–319.
Becker, Kelly. 2012. "Methods and How to Individuate Them", in K. Becker and T. Black (eds.), *The Sensitivity Principle in Epistemology*, New York: Cambridge University Press.
Berend, Daniel, and Jacob Paroush. 1998. "When Is Condorcet's Jury Theorem Valid?", *Social Choice and Welfare* 15: 481–88.
Berend, Daniel, and Luba Sapir. 2005. "Monotonicity in Condorcet Jury Theorem", *Social Choice and Welfare* 24: 83–92.
Boland, Philip J. 1989. "Majority Systems and the Condorcet Jury Theorem", *Journal of the Royal Statistical Society*. Series D (The Statistician) 38: 181–89.
Burgman, M. A. 2005. *Risks and decisions for conservation and environmental management.* Cambridge: Cambridge University Press.
Christensen, David. 2007. "Epistemology of Disagreement: The Good News", *The Philosophical Review* 116: 187–217.
-------. 2009. "Disagreement as Evidence: The Epistemology of Controversy", *Philosophy Compass* 4/5: 756–67.
-------. 2010. "Higher-Order Evidence", *Philosophy and Phenomenological Research* 81: 185–215.

Clarke-Doane, Justin. Forthcoming. *Morality and Mathematics*, Oxford University Press.

Clemen, R. T., and R. L. Winkler. 1999. "Combining probability distributions from experts in risk analysis", *Risk Analysis* 19: 187–203.

Condorcet De, Nicolas. 1785. *Essai Sur L'Application de L'Analyse à La Probabilité Des Décisions Rendues à La pluralité des voix.*

Dietrich, Franz. 2008. "The Premises of Condorcet's Jury Theorem Are Not Simultaneously Justified", *Episteme* 5: 56–73.

Dietrich, Franz, and Christian List. 2004. "A Model of Jury Decisions Where All Jurors Have the Same Evidence". *Synthese* 42, 175–202.

Elga, Adam. 2007. "Reflection and Disagreement", *Noûs* 41: 478–502.

Enoch, David. 2010a. "Not Just a Truthometer: Taking Oneself Seriously (but not too Seriously) in Cases of Peer Disagreement", *Mind* 119: 953–97.

-------. 2010b. "The Epistemological Challenge to Metanormative Realism: How Best to Understand It and How to Cope with It", *Philosophical Studies* 148: 413–38.

Estlund, D., Waldron, B. Grofman, and S. Feld. 1989. "Democratic Theory and the Public Interest: Condorcet and Rousseau Revisited." *American Political Science Review* 83, 1317–40.

Estlund, David, 1994. "Opinion leaders, independence and Condorcet's Jury Theorem", *Theory and Decision* 36, 131–62.

Feldman, Richard. 2006. "Epistemological Puzzles about Disagreement", in S. Hetherington (ed.), *Epistemology Futures*, Oxford: Oxford University Press.

Fogal, Daniel. 2016. "Reasons, Reason, and Context", in E. Lord and B. Maguire, *Weighing Reasons*, New York: Oxford University Press.

-------. 2018. "On the Scope, Jurisdiction, and Application of Rationality and the Law", *Problema* 12: 21–57.

-------. Forthcoming. "Rational Requirements and the Primacy of Pressure", *Mind*, https://doi.org/10.1093/mind/fzz038.

Fogal, Daniel, and Olle Risberg. Forthcoming. "The Metaphysics of Moral Explanations", in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 15*.

Goldman, Alvin. 2001. "Experts: Which Ones Should You Trust?", *Philosophy and Phenomenological Research* 63: 85–110.

Grofman, Bernard, Guillermo Owen, and Scott L. Feld. 1983. "Thirteen Theorems in Search of the Truth", *Theory and Decision* 15: 261–78.

Hedden, Brian. 2017. "Should Juries Deliberate?", *Social Epistemology* 31: 368–86.

Horowitz, Sophie. 2014. "Epistemic Akrasia", *Noûs* 48: 718–44.

Joyce, Richard. 2006. *The Evolution of Morality*, Cambridge: MIT Press.

Kaniovski, Serguei. 2010. "Aggregation of Correlated Votes and Condorcet's Jury Theorem", *Theory and Decision* 69: 453–68.

Kelly, Thomas. 2005. "The Epistemic Significance of Disagreement", in J. Hawthorne and T. Gendler (eds.), *Oxford Studies in Epistemology, vol. 1*, Oxford: Oxford University Press.

-------. 2011. "Peer Disagreement and Higher Order Evidence", in A. Goldman and D. Whitcomb (eds.), *Social Epistemology: Essential Readings*, Oxford: Oxford University Press.

Kornblith, Hilary. 2010. "Belief in the Face of Controversy," in R. Feldman and T. A. Warfield (eds.), *Disagreement,* Oxford: Oxford University Press.

Lackey, Jennifer, and Ernest Sosa (eds.). 2006. *The Epistemology of Testimony*, Oxford: Oxford University Press.

Lasonen-Aarnio, Maria. 2014. "Higher-Order Evidence and the Limits of Defeat", *Philosophy and Phenomenological Research* 88: 314–45.

Lehrer, Keith, and Carl Wagner. 1981. *Rational Consensus in Science and Society,* Dordrecht: Reidel.

List, Christian, and Robert E. Goodin. 2001. "Epistemic Democracy: Generalizing the Condorcet Jury Theorem", *Journal of Political Philosophy* 9: 277–306.

Lutz, Matt. 2019. "Explanationism provides the best explanation of the epistemic significance of peer disagreement", *Philosophical Studies*, https://doi.org /10.1007/s11098-019-01286-0

Mill, John Stuart. 1859. *On Liberty*. In *On Liberty and the Subjection of Women*, 1879. New York: Henry Holt & Co.

Mulligan, Thomas. forthcoming. "The Epistemology of Disagreement: Why not Bayesianism?", *Episteme*. doi:10.1017/epi.2019.28

Owen, Guillermo. 1986. "'Fair' Indirect Majority Rules." In B. Grofman and G. Owen (eds.), *Information Pooling and Group Decision Making*, Greenwich, CT: JAI Press.

Pivato, Marcus. 2013. "Voting Rules as Statistical Estimators", *Social Choice and Welfare* 40: 581–630.

Pollock, John. 1986. *Contemporary Theories of Knowledge*, Totowa, NJ: Roman and Littlefield.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Riegler, Alexander, and Igor Douven. 2009. "Extending the Hegselmann–Krause model III: From single beliefs to complex belief states", *Episteme* 6: 145–63.

Risberg, Olle, and Folke Tersman. 2019. "A New Route from Moral Disagreement to Moral Skepticism", *Journal of the American Philosophical Association* 5: 189–207.

-------. 2020. "Disagreement, Indirect Defeat, and Higher-Order Evidence". In M. Klenk (ed.), *Higher-Order Evidence and Moral Epistemology*. Routledge.

Rowbottom, Darrel. 2018. "What is (Dis)Agreement?", *Philosophy and Phenomenological Research* 97: 223–36.

Schupbach, Jonah. 2018. "Robustness Analysis as Explanatory Reasoning", *The British Journal for the Philosophy of Science* 69: 275–300.

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value", *Philosophical Studies* 127: 109–66.

Tersman, Folke. 2017. "Debunking and Disagreement", *Noûs* 51: 754–74.

Vavova, Katia. 2018. "Irrelevant Influences", *Philosophy and Phenomenological Research* 96: 134–52.

Weisberg, Michael. 2006. "Robustness Analysis", *Philosophy of Science* 73: 730–42.

White, Roger. 2010. "You Just Believe That Because…", *Philosophical Perspectives* 24: 573–615.

Wimsatt, William. 2011. "Robust Re-engineering: A Philosophical Account?", *Biology and Philosophy* 26: 295–303.

Woodward, Jim. 2006. "Some Varieties of Robustness", *Journal of Economic Methodology* 13: 219–40-

Worsnip, Alex. 2018. "The Conflict of Evidence and Coherence", *Philosophy and Phenomenological Research* 96: 3–44.

Zamora Bonilla, Jesus. 2012. "The economics of scientific knowledge", in U. Mäki (ed.), *Philosophy of Economics*, Amsterdam: Elsevier.

Zhao, Haicheng. 2018. "Knowledge Without Safety", *Synthese*. https://doi.org/10.1007/s11229-018-1881-x