A role for judgment aggregation in coauthoring scientific papers

Erkenn (2018) 83:231–252
https://doi.org/10.1007/s10670-017-9887-1

ORIGINAL RESEARCH

A Role for Judgment Aggregation in Coauthoring Scientific Papers

Liam Kofi Bright[1] · **Haixin Dang**[2] · **Remco Heesen**[3]

Received: 18 April 2016 / Accepted: 15 February 2017 / Published online: 23 March 2017

Abstract This paper addresses the problem of judgment aggregation in science. How should scientists decide which propositions to assert in a collaborative document? We distinguish the question of what to write in a collaborative document from the question of collective belief. We argue that recent objections to the application of the formal literature on judgment aggregation to the problem of judgment aggregation in science apply to the latter, not the former question. The formal literature has introduced various desiderata for an aggregation procedure. Proposition-wise majority voting emerges as a procedure that satisfies all desiderata which represent norms of science. An interesting consequence is that not all collaborating scientists need to endorse every proposition asserted in a collaborative document.

Author order is alphabetical. The authors are happy to confirm that every proposition asserted here is supported by at least two authors, consistent with our main thesis.

✉ Remco Heesen
rdh51@cam.ac.uk

Liam Kofi Bright
lbright@andrew.cmu.edu

Haixin Dang
had27@pitt.edu

[1] Department of Philosophy, Baker Hall 161, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

[2] Department of History and Philosophy of Science, University of Pittsburgh, 1017 Cathedral of Learning, Pittsburgh, PA 15260, USA

[3] Faculty of Philosophy, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK

1 Introduction

Suppose a group of inquirers wants to announce the results of their research to the world: how should they decide what they declare to be the results of their investigation? Resolving this question is especially relevant as today the vast majority of articles found in peer-reviewed scientific journals are authored by multiple researchers (King 2013). How should a group of collaborating scientists come to agree about what their collaboration will report as their results?

The literature on judgment aggregation provides a suggestive formal model for studying this question. However, some philosophers have expressed concerns that this model is unable to adequately capture the complexities of social interactions in science, especially deliberative practices (Magnus 2013; Wray 2014). In addition, List and Pettit (2002) have proven an impossibility theorem which denies the existence of an aggregation procedure which satisfies universal domain, anonymity and systematicity; features of a judgment aggregation rule that may seem appealing in the context of scientific collaboration. These difficulties appear to have made the application of judgment aggregation to scientific collaborations unattractive. In this paper, we argue against these criticisms of judgment aggregation and in favor of proposition-wise majority voting as the most appropriate aggregation procedure for collaborating scientists.

In Sect. 2, we argue that the question of group belief should be distinguished from the question of what is reported in a published unit. By published units, we mean things like articles, papers, books, presentations, online preprints—any form of statement addressed to the scientific community. We claim that the philosophical objections regarding the kind of aggregation procedures suggested by formal models of judgment aggregation apply to the question of group belief rather than to the question of what to report, which is of independent interest from the perspective of social epistemology. Opponents and proponents of judgment aggregation have both assumed that aggregation functions are models of group belief. We use aggregation functions as models of collective reporting, which avoids the problems of group belief.

In Sect. 3, we clarify the role of deliberation in collaborative practice and its relationship to judgment aggregation. We argue that the voting procedures that results from a judgment aggregation function represent a standard that successful deliberation must meet, but the voting procedures themselves do not replace deliberation. Section 4 discusses how we can account for the role of values in judgment aggregation.

In Sects. 5 and 6 we advance our normative proposal for judgment aggregation in science. We connect the norms currently endorsed by working scientists with results in judgment aggregation and show that the former can be simultaneously satisfied. We argue that proposition-wise majority voting can in fact be a reasonable aggregation procedure for scientists to adopt when deciding what to report in a published unit (within a suitably restricted domain). We show that some plausible alternative proposals fall victim to the impossibility theorems of List and Pettit and subsequent authors, while our proposal does not.

Section 7 gives more details on what we take our normative recommendations to be for two specific applications. We first consider the case of consensus conferences, a type of conference designed specifically to elucidate the scientific consensus on a small set of interrelated questions. Consensus conferences have been particularly popular in the field of medicine, although this popularity has recently waned (Solomon 2015). The second case we consider is that with which this introduction began: the problem of collaborative publications.

2 Collective Epistemology and Judgment Aggregation

Collective epistemology is concerned with the study of the knowledge possessed by social groups. There exist two main approaches to thinking about group knowledge: summativism and non-summativism. The summativist position is that all group phenomena can be understood entirely in terms of individual phenomena. For example, a summativist may claim that what a group believes is simply what all or the majority of individuals in the group believe. Summativism has been challenged by divergence arguments, which purport to describe cases when individual beliefs diverge from the group belief. For example, the group G may believe that p, even if none of the individual members of G believe that p. These divergence cases have been taken to support a non-summativist position. Non-summativists argue that a group is an epistemic subject in its own right and the group belief is distinct from the individual beliefs. A notable non-summativist position has been defended by Gilbert (1987, 1992), who has argued for a joint acceptance account of group knowledge. Under this account, a group is said to believe that p if and only if all or most members of the group express willingness to accept that p as the view of the group regardless of whether they personally believe that p, and under conditions of common knowledge that others in the group are so committing.

In this paper, we do not take a position either for or against summativism. In fact, our positive thesis is compatible with either a summativist or a non-summativist account of group belief. The debate between summativism and non-summativism has obscured what is at stake when determining the epistemic status of a published piece of work. The target of our discussion is the published unit itself: a set of propositions which are (logically) related to each other. How can a group of collaborating scientists come to consensus about the set of propositions to be reported in a published unit?

Most philosophers have taken a coauthored paper to express the belief of the group. Here, we take the position that the published unit is analytically separate from the group belief. A published unit is a public document containing sets of logically related propositions. Collective belief and the published units produced by that collective cannot be assumed to be identical. This distinction has mostly been overlooked in the literature.

A partial exception is Huebner et al. (2017) which treats collective authorship as a separate issue from collective action or group knowledge. This, however, is directed towards a different end than our project. Huebner et al. are interested in methods for ensuring accountable authorship in situations of radical collaboration,

which are situations that make it difficult to keep track of who contributed what to a research project. We, on the other hand, are interested in strategies for how a group may come to write a single report together, which need not involve any difficult questions of attributing responsibility to coauthors.

The major reason for making the distinction between collective belief and published unit is that groups may come to them through different processes. A purely summative belief formation may simply involve the majority of the group coming to believe that *p*. The example of a non-summative group belief in Gilbert (1987) concerns a group of people discussing a poem together. The non-summative group belief, a certain interpretation of the poem, is established through a deliberation and dialoguing process. Fagan (2011) argues for a middle way position where the group belief is established through the interaction between individual members' beliefs; she calls this third option interactive belief. But these processes of collective belief formation do not necessarily describe the construction of a *written* document. Wagenknecht (2015) observes that research groups will often have hierarchical authorship practices, where one or a couple of members are responsible for writing the final report for publication; the other members may not contribute to writing the paper at all, even though they may have been integral to the genesis and development of the ideas presented in the paper. These hierarchical authorship practices result in a final report which is not *necessarily* the group belief as described by the summative, non-summative, or interactive accounts of group belief formation.

Some philosophers argue that groups cannot be bearers of beliefs, instead groups can only accept claims. Acceptance of a proposition is a much weaker state than belief. As Cohen has put it, "[t]o accept that *p*, is to have or adopt a policy of deeming, positing, or postulating that *p*—i.e., of including that proposition or rule among one's premises for deciding what to do or think in a particular context, whether or not one feels it to be true that *p*" (Cohen 1992, p. 4). Wray (2001) has argued that 'collective beliefs' are not species of belief proper but rather a species of acceptance. This is because proper beliefs are involuntary and difficult to change (according to many epistemologists; see Schwitzgebel 2015, Sect. 2.5, and references therein). According to Wray, groups fail to have proper beliefs because they often adopt beliefs for realizing practical goals, i.e., voluntarily. Groups are also prone to change their views, often for irrelevant reasons, whereas proper beliefs are more stable over time.

Perhaps such philosophers could say that while a collaborative paper may be separate from a collective belief, the paper can be a set of collectively accepted propositions. We would have no quarrel with this reading of our proposal: a model of group acceptance could be made consistent with our view on paper construction in science. We do not specify one particular theory of group acceptance in this paper, since the relationships between the various individual attitudes and collective attitudes is complex and it would take us too far afield to fully delve into the matter (see List 2014). The important point for us is that we can separate analyses of collective belief formation from analyses of paper construction. We will argue throughout this paper that some of the current skepticism about judgment

aggregation when applied to science is the result of a focus on collective belief formation instead of weaker attitudes like acceptance and collective reporting.

On the other hand, there are some philosophers who do not outright reject applying judgment aggregation to science (see Solomon 2006; Wray 2014). Rolin (2015) is sympathetic to the idea that judgment aggregation can be applied to science. She argues that judgment aggregation procedures can help individuals in a research group arrive upon a consistent set of views. She believes that members of a group ought to jointly commit to a judgment aggregation procedure in order to "collectivize reason". Agreement among members on a judgment aggregation procedure is key to the group maintaining internal consistency of their views and reasons. However, she does not describe what such a procedure would be. What kind of judgment aggregation should scientific research groups be committed to in order to maintain internal consistency? The solution we offer in Sect. 6 is exactly such a procedure.

3 The Role of Deliberation

One of the common objections to judgment aggregation approaches is a concern that voting procedures will replace deliberation. To put this objection starkly, judgment aggregation results may be read as suggesting that scientists are not allowed to talk to each other. Instead they are merely to submit their opinions to a black box which then mechanically aggregates them into a published unit. Magnus, for example, writes that aggregation procedures treat scientists "merely as separate inputs to an algorithm" (Magnus 2013, p. 847).

The worry that deliberation may be supplanted by judgment aggregation has been discussed by Wray (2014), writing in response to Solomon (2006). Solomon has argued that groups in deliberation are prone to groupthink; for example, members of a group may be peer pressured into consensus and may suppress relevant evidence when facing disagreement. She concludes that judgment aggregation, *without deliberation*, avoids groupthink and results in a collective view that is in accordance with all the available evidence. Wray argues that Solomon was too quick to dismiss deliberation. He claims that in the context of collaborative teams writing a coauthored paper *consensus* is necessary to achieve a group view and therefore, deliberation remains an important process. He argues that there are strategies to avoid groupthink at different points of a collaborative process. Wray does not completely reject judgment aggregation, but he thinks it is only sometimes relevant for collaborative groups.

We agree with Wray that judgment aggregation ought not to replace deliberation altogether. Collaborative scientific work virtually always involves extensive deliberation among the collaborators about the results, the implications, and many other aspects of their work. We do *not* believe that scientists should stop their current deliberative practices in favor of a voting procedure.

But we disagree with Wray that consensus (read as unanimity) is necessary if a group is to publish the results of their research as a coauthored article. Here our distinction between collective belief and collective reporting does important work.

Some sort of consensus *may* be necessary for collective belief, but consensus is not at all necessary for collective reporting.

The dialectic between our view and that of Wray is subtle. Wray adopts a joint acceptance account for collective reporting: "Research teams need to deliberate in order to reach a consensus about what view will stand as the view of the group" (Wray 2014, p. 292). Elsewhere, Wray (2001) has argued for a distinction between collective acceptance and collective belief; he argues that they pick out different cognitive states. So for Wray, collective reporting is a kind of collective acceptance, arrived at through a consensus or joint acceptance. Here we argue for an even further distinction: collective reporting, separate from collective acceptance, does not require consensus. We distinguish ourselves from Wray's account for collective reporting because we believe consensus is too strong a requirement in the context of collaborative paper writing. Parts of a paper may rely on special expertise of a few members which cannot be assessed by the other members, especially in large-scale scientific projects with hundreds or thousands of collaborators, e.g., in high-energy physics. In these cases, joint acceptance or consensus is not possible, and so it cannot be the goal of deliberation.

In our view, the desired outcome of deliberation is to reach *sufficient* agreement among the coauthors about what results to collectively report. Sufficient agreement is reached when the individual collaborators' views are such that when an appropriate judgment aggregation function is applied to them, the resulting published unit is consistent. We will argue below that proposition-wise majority voting is this appropriate judgment aggregation function. The voting process that is highlighted in judgment aggregation is hypothetical, and only comes into play after any and all deliberation the scientists care to do has taken place. We recognize that deliberation is an important part of the scientific process and we do not think a voting procedure will replace it or be a better alternative.

Our judgment aggregation method says very little about *how* to best deliberate. Instead, the formal requirements of judgment aggregation set a standard for the *result* of deliberation.

4 The Role of Values

Another objection to judgment aggregation approaches is a concern about inductive risk. In particular, Magnus (2013) worries that judgment aggregation procedures idealize scientific reasoning such that the inference patterns captured by these procedures cannot account for value judgments that scientists inevitably must make.

Scientists typically make inferences from data to conclusions, or at least decide whether to publish the claim that some evidence sufficiently supports a hypothesis. Such inferences can be uncertain, and to make them one must make the kind of value judgments discussed by Rudner (1953) and Douglas (2009). Inferences of the form $E \rightarrow H$ can thus only be assessed if we know what risks are involved in inferring H from E. Magnus' worry is "[i]f we merely poll scientists [on $E \rightarrow H$], then we will be accepting whatever judgments accord with their unstated values" (Magnus 2013, p. 847). As judgment aggregation functions "elide the role of

values" (p. 848), they are not an adequate model of scientists' collective knowledge.

In Magnus (2013), this objection from inductive risk is aimed at the level of the scientific community, in cases where it is important to base policy decisions on what the community knows. However, this objection can be extended to the level of research groups. If the research group is reporting research that has policy ramifications, how can their value judgments be accounted for?

This objection is related to how deliberation relates to judgment aggregation. Values may be elided if deliberation does not take place. As we have argued earlier, deliberation and judgment aggregation are not mutually exclusive. Thus, judgment aggregation will not replace the kind of analytic-deliberative process advocated for by Douglas (2009) to deal with risks and values in science.

We argue below that the outcome of the analytic-deliberative process must be consistent with a judgment aggregation function, reflecting a minimum level of agreement between the scientists at the end of the process. If value judgments are explicitly discussed, these value judgments should be made part of the agenda to be (hypothetically) voted on. That is, in settling $E \rightarrow H$, various judgments about the risk of the inference should be voted on as well. This vote is hypothetical; through deliberation, we may find that disagreement over the risk is widespread and therefore an aggregation is not possible (we will outline conditions for this below). Or we may find that there is enough agreement among scientists about what the risks are, making it possible to settle $E \rightarrow H$. Judgment aggregation does not need to elide the role of values.

Someone sympathetic to Magnus' objection might respond that whatever norms governed the deliberation will themselves be elided in publication. However, we think that to the extent that this is eliding value judgements it represents an inevitable feature of publication. Suppose that a paper were single-authored, and thus could be thought of in judgement aggregation terms as being subject to a dictator. The dictator still has to decide whether conclusions could be safely inferred from the available evidence, based on her values. While some argument for her decisions could be given in the paper, she would quickly faces a regress if she attempted to explicitly note all of the value judgements that informed her decisions. Once propositions stating the relevant value judgement are affirmed in the paper they become subject to the same requirements of evidence and support as any other scientific assertion, requiring further value judgments to justify their inclusion. In eliding the value judgements that inform deliberation coauthorship teams are in no worse a position than the dictator. Some value judgements that informed paper construction are inevitably elided in any document of finite length, whether the paper is coauthored or not.

We suspect that part of what might make somebody think this is a special problem for using judgement aggregation procedures to get at group opinion is that one has in mind group belief in one of the stronger senses discussed in Sect. 2. Under a summativist position on group belief $E \rightarrow H$ should only be affirmed if all or most of the group believe it. However, due to differing risk thresholds it could come about that inconsistent standards for 'belief' are applied by various members of the group, and the unity revealed by a vote is merely illusory. Perhaps this creates

difficulty for a summativist wishing to say that the group believes $E \rightarrow H$. Under a non-summativist position perhaps the proper procedure for forming group beliefs must necessarily involve negotiation over the risks involved in affirming $E \rightarrow H$. A group that fails to go through this process may thereby fail to form a properly constituted group belief in $E \rightarrow H$, even if they come to agreement on the proposition via vote.

We concede both these points as applied to group belief. However, we are concerned in this paper with the formulation of a collaborative document, whose content need not be the same as a group belief. As the case of the dictator makes clear: some decisions must be made as to what goes into a paper, decisions that will necessarily involve eliding or rendering non-apparent some of the value judgements that inform them. As such, we do not think the objections raised here constitute a problem for collaborative document formation, when that is considered separately from group belief formation.

5 Desiderata for Judgment Aggregation in Collaborative Science

How, if at all, does the formal theory of judgment aggregation yield a solution to the problem we have sketched?

One approach would be to investigate directly the epistemic properties of judgment aggregation functions, e.g., their truth-conduciveness. But to recommend the outcomes of such an investigation to scientists would require careful consideration of its consequences at the social level. Recent work in social epistemology has shown that the epistemically desirable features of individual scientists' (here, individual collaborations') behavior do not necessarily scale up to epistemically desirable features at the social level (Mayo-Wilson et al. 2011). To our knowledge the question of what epistemic features individual papers need to have to optimally contribute to science has not been settled, and we do not have the ambition to settle it here. So it is not clear what we would be looking for if we studied the epistemic properties of judgment aggregation functions directly.

For this reason we take a more indirect approach. We take the existing norms of scientific publishing as given. For the purposes of this paper we assume that the existing norms lead to fairly good (if not necessarily optimal) epistemic outcomes at the social level, at least for single-authored publications. We also assume that coauthored published units are held to the same normative standards as single-authored publications. We then investigate the question which judgment aggregation function collaborating scientists should use in order to satisfy these existing norms.

This approach has a number of advantages. First, rather than having to deduce the desiderata for our judgment aggregation function from a substantive social epistemological argument, we can obtain them inductively by studying the norms scientists actually take themselves to be held to. The latter is a well-studied subject, and we can draw upon existing work in the philosophy and sociology of science. Second, the existing norms lend themselves well to being translated into desiderata on judgment aggregation functions that have already been studied extensively. In

contrast, only a few papers have studied the epistemic properties of judgment aggregation functions directly (these papers draw on the Condorcet jury theorem in various ways, see List 2005; Hartmann and Sprenger 2012; Bozbay et al. 2014, and references therein). Third, as we will argue, the existing norms of science yield a surprisingly specific verdict on the question which judgment aggregation function should be used. And finally, it will be easier to convince scientists that they should aggregate their judgments in a particular way if this requirement is seen to follow from norms they already take themselves to be committed to.

We shall base our arguments on the norms of science studied by Merton (1942). That is, scientists should impersonally assess claims (universalism), freely share information (communism), be motivated by more than mere personal gain (disinterestedness), and subject claims to rigorous criticism (organized skepticism).

The Mertonian norms have been widely discussed and criticized by other sociologists of science. It goes beyond the scope of the present paper to survey the entire controversy, from the existence of counter-norms to additional norms, or the extent to which these norms are part of scientific practice (see Merton 1942; Mulkay 1976; Gibbs 1981). Nevertheless, the Mertonian norms persisted. Anderson et al. (2010) have found continuing wide support for these norms through a large scale sociological study (including focus groups and a survey of scientists in different fields). The norms may present an overly idealistic and simplistic picture of science, but it is still one that scientists take themselves to be commited to. Our arguments depend on this commitment rather than on whether the norms are honored in practice.

Merton's and Anderson et al.'s work is based on study of the scientific community at large, rather than specifically on research on groups as sub-units of the scientific community. A potential concern is that the largely egalitarian Mertonian norms only apply at the level of the community but not at the level of collaborating groups; after all, even if scientists are equals as members of the scientific community, when collaborating on a specific project they each have specific roles and there is often a clear hierarchy among them.

We do not think that this is a tenable view of the operation of scientific norms. A huge part of scientists' interaction with each other *qua* scientists happens at the level of research group; labs, departments, and collaborative groups more generally, are a mainstay of institutional and group life for scientists. If these norms are not operative in the context of such interactions then it is difficult to see why scientists would think of them as general norms of scientific life. Further, role differentiation and hierarchy exist in the context of the broader scientific community as well as just within lab groups. If the Mertonian norms were not applicable when role differentiation and hierarchy are present, they would scarce be operative within scientific communal life at all. We hence take it that scientists themselves would agree that these norms should guide their own practice in the context of collaborative research.

A judgment aggregation function takes the judgments of a group of individuals on a number of logically related propositions as input and yields the propositions to be asserted in the published unit as output. The set of propositions on which the individuals give their judgment is called the agenda. We assume that at least two

propositions and either their conjunction or disjunction are included, and that for any included proposition its negation is also included. We assume that individual judgments are complete and consistent (in the sense given below).

We now introduce a number of constraints one might place on a judgment aggregation function. For each of them, we discuss whether existing norms of scientific publishing support this constraint.

1. *Completeness*: The aggregation function judges all relevant propositions, i.e., for every proposition on the agenda, the published unit asserts it or its negation.

Completeness is not generally supported by a norm of science. It is true that writing on some topics requires saying something about related topics, but this is never so specific as to require either asserting or denying specific propositions.

2. *Consistency*: The published unit is logically consistent.

Consistency reflects the norm that a published unit should not contradict itself. That this is a norm we take to be evident. Self-contradiction is never acceptable in a paper, be it single-authored or coauthored.

The normative status of consistency may be contested by those who note that high level theories in physics are known to contradict each other and yet the scientific community still seems happy to endorse them (Priest 2006, chapter 9). But such a purported counterexample fails to pay attention to the level of analysis we are working at. It may well turn out that the belief set we are led to by aggregating multiple published units is itself inconsistent. In this paper, however, we are concerned with those judgment aggregation procedures that operate to produce published units. Such units are expected to be consistent.

3. *Deductive closure*: Any proposition that logically follows from those asserted in the published unit (and is included in the agenda) should be included in the published unit.

Failure of this desideratum would be a case where a collaborative team asserts, e.g., p and $p \rightarrow q$ but fails to assert q. Since one may legitimately criticize a paper for its logical consequences regardless of whether it is single-authored or collaboratively authored, it seems that the norms of scientific publishing imply that collaborative teams are committed to the logical consequences of their assertions. The desideratum of deductive closure formalizes this, requiring logical consequences to be explicitly asserted if they are part of the agenda.

If completeness is required, consistency entails deductive closure. But since we will consider the possibility of not requiring completeness, it is important to distinguish these two desiderata. Consider again the team asserting p and $p \rightarrow q$, and suppose that q and $\neg q$ are included in the agenda. Deductive closure requires that the team also asserts q, whereas consistency forbids the team from also asserting $\neg q$. Hence, it is possible to be deductively closed but inconsistent (assert both q and $\neg q$) or to be consistent but not deductively closed (assert neither q nor $\neg q$).

4. *Unanimity preservation*: Whenever all individuals assert a given proposition, the published unit should assert it too.

Unanimity constitutes the ideal case. If a collaborative team is going to create a published unit, at the very least it should include those propositions (among those that are relevant, as determined by the agenda) about which they all agree. Further, as we shall discuss below, unanimity preservation has some support in the practice of science.

5. *Anonymity*: All individuals are considered equal. More formally, if the judgments of two individuals are switched, the published unit does not change.

This is a reflection of the universalism norm (Merton 1942). This norm requires that the contributions individual scientists make are to be evaluated independently of the scientist who contributed them. In science it is the arguments that count, not the personalities, though here we stress that these norms are not necessarily always honored in practice! Recent evidence confirms that scientists remain committed to this idea (Anderson et al. 2010, p. 9 and p. 15).

This norm entails that once the evidence has been gathered, decisions about what should be included in the published unit should satisfy anonymity. To see why this is so, it is important to pay attention to the phenomenon we intend to model. Those involved in the collaboration have shared ideas and arguments; what remains is to decide on what to report to the community at large. To be consistent with universalism, whatever influence the personalities of the scientists involved may permissibly have can come only through their presentation of arguments and reasons that persuade their fellows. The mere fact that a particular person has a particular preference cannot be an additional reason. To think otherwise is to evaluate the object of preference in part as a function of the particular person expressing the preference, a direct violation of universalism.

A possible objection to anonymity claims that the opinion of an expert on a particular topic is often given greater weight. We discuss this objection below under "expert rights". We argue there that expert rights are not supported by the norms of science, thus providing further (indirect) support for anonymity.

6. *Systematicity*: If the judgments of all individuals are the same on two propositions, then the published unit should either assert both of them or deny both of them.

We will defend systematicity as a desideratum by linking it to organized skepticism, another norm of science identified by Merton (1942). According to this norm scientists should treat all propositions or commitments as open to investigation, not to be prejudged or accepted without being submitted to the reasoned evaluation of the scientific community. Once again, both focus group study and a national survey reported support for organized skepticism (Anderson et al. 2010, p. 9 and p. 15).

We interpret this norm as meaning that in the case of individual published units all propositions should be held to the same standards regarding how much support

they need from participants before they are asserted. That is, we take systematicity to interpret organized skepticism in the context of collaborative published units. The argument for this point is precisely parallel to the argument for anonymity via universalism. Whatever preference is to be expressed for a particular proposition must be a result of the reasons or evidence that have been presented in its favor. Any additional preference shown to a proposition would constitute a violation of organized skepticism.

One may contest this norm on Kuhnian grounds. Part of being inculcated into a scientific tradition involves learning to take some things as not up for debate, and there may well be good reasons for this dogmatism (Kuhn 1977). This observation, however, is consistent with believing in systematicity as a norm of scientific publishing. For such paradigm-defining propositions will receive universal assent from the collaborators. So as long as unanimity preservation is required scientists need not be willing to abandon paradigm-defining propositions. Hence we can insist that there are no propositions that require more (or less) support from the collaborative team in order to make it into the published unit, without contradicting Kuhn's insight.

While we take systematicity to be a norm of scientific publishing, the following two departures from systematicity merit explicit attention. First, one might like to distinguish between propositions that represent conclusions of the paper and propositions that represent reasons or evidence for those conclusions. This suggests judgment aggregation functions that violate systematicity. On a reasons-based approach, one might first determine what the published unit is to say about the reasons, and then determine what the published unit says about the conclusions using deductive closure (disregarding individual judgments on the conclusions). Or, on a conclusions-based approach, one might first determine which conclusions to assert, and allow this to influence which reasons are asserted. There has been some philosophical discussion of these approaches (Magnus 2013).

We think these approaches do not yield good recommendations for judgment aggregation in science. First, they require a neat separation of propositions into conclusions and reasons, which may not exist in practice: how about intermediate conclusions, for instance? Second, in order for these approaches to have any advantages over systematic ones (in the sense of avoiding the impossibility theorems we discuss below), either the reasons or the conclusions must be logically independent. Again, we think this will often not be the case in practice.

We admit that if the case can be made that these conditions are satisfied in a particular judgment aggregation problem, reasons- or conclusions-based approaches may be reasonable. But our aim here is to give a recommendation that applies more generally. So we set these approaches aside.

A second departure from systematicity (as well as anonymity) might be motivated by a desire to give special consideration to the opinion of experts on particular propositions. The Nobel prize winning physicist Carlo Rubbia, for one, seemed to think that there was some basis for doing this. When discussing with his research team whether to include some contentious claims in an upcoming presentation he said: "I cannot neglect the fact that people who are working on [those aspects of the experiment under discussion] have more weight than people

who aren't" (Taubes 1986, p. 218). More generally, in contemporary big science it is unavoidable that some scientists have a better understanding of some parts of the collaborative project than others. Should this not be reflected in the judgment aggregation function? For this purpose we introduce the idea of expert rights.

7. *Expert rights*: A scientist has expert rights on a proposition if the published unit always follows that scientist's judgment on that proposition and its negation.

We do not think expert rights reflect a norm of scientific publishing. We think, rather, that expertise should play its role in deliberation. One important part of being an expert is to "be able to 'give an account' of what it is that she is expert in" (Annas 2001, p. 244). An expert should be able to explain why the other scientists should agree with her on the propositions that she is expert in. As a result, the other scientists will come to believe what the expert believes either through discussion and deliberation or because they simply defer to the expert's expertise. The expert opinion will then make it into the published unit by unanimity preservation.

If one or more scientists, for whatever reason, sufficiently doubt the expert's opinion to maintain an individual judgment different from hers, then the mere fact that she is the expert should not make her opinion prevail. It may prevail in the end, but in our view it should do so on the basis of a systematic and anonymous aggregation. Hence we think anonymity and systematicity reflect norms of scientific publishing, and expert rights do not. A related reason to reject expert rights will be given in Sect. 6.

8. *Acceptance/rejection neutrality*: Whenever all individuals flip their judgment on a given proposition, the published unit flips as well.

Acceptance/rejection neutrality is essentially systematicity applied to a proposition and its negation. As a result we think that acceptance/rejection neutrality must be satisfied to respect Merton's norm of organized skepticism. Doing otherwise would be a way of favoring a proposition over its negation (or vice versa) not on the basis of reasons or evidence, in violation of the norm. From a formal perspective this is a fairly strong requirement: we will see below that if it is accepted our proposal is in an important sense the only possibility.

9. *Universal domain*: The aggregation function yields a published unit for any possible combination of (complete and consistent) individual judgments.

Universal domain requires that a published unit is produced regardless of the views of the individuals involved, i.e., no matter how much they disagree. This does not seem to be a norm of scientific publishing: in cases of widespread disagreement no published unit may be produced at all. Alternatively, some scientists may take their name off a paper if they find themselves unable to support its conclusions. The latter is commonly accepted practice, especially in large collaborations. This happened, for example, when the Collision Detector at Fermilab (CDF) produced evidence that some interpreted as evidence for mysterious extra muons: ghost particles that are suggestive of new physics. The results were published in an online

preprint, but they were so controversial that nearly a third of the roughly 600 scientists involved refused to sign it (CDF Collaboration 2008).

6 Possibilities and Impossibilities in Judgment Aggregation

In the previous section we argued that a judgment aggregation function for collective reporting should satisfy consistency, deductive closure, unanimity preservation, anonymity, systematicity, and acceptance/rejection neutrality, while completeness, expert rights, and universal domain are not required. Now it is time to consider whether the former can be simultaneously satisfied. Our starting point for this discussion is a result by List and Pettit (2002). They prove that there exists no judgment aggregation function which satisfies completeness, consistency (and hence deductive closure), anonymity, systematicity, and universal domain (List and Pettit 2002, Theorem 1).

It is perhaps disappointing that no such function exists, but it is not a major setback. After all, we have argued that neither completeness nor universal domain reflect norms of scientific publishing. We can drop either of those to avoid impossibility. We argue that the most promising solution is found by dropping universal domain. Part of our argument consists in showing that the two most plausible alternatives that suggest themselves if we retain universal domain do not get us very far, so we first consider these alternatives.

The first alternative is to drop the requirements of anonymity and systematicity, replacing them with expert rights. However, we quickly run into impossibility again. Say that two agenda propositions p and q are conditionally dependent if there exists a set of agenda propositions such that the propositions in that set are inconsistent with p and q but consistent with p or q. Suppose there are two scientists, each of which has expert rights on one of a pair of conditionally dependent propositions (this may not always be the case, but a sufficiently general norm for collaborative publishing would have to cover this case). Then there is no judgment aggregation function which satisfies consistency, unanimity preservation, expert rights, and universal domain (Dietrich and List 2008a, Sect. 3.1).

Dietrich and List (2008a) consider ways to avoid this impossibility by weakening universal domain. This would take us to a situation where both anonymity and universal domain are violated. However, we have already argued that anonymity is a norm of science, reflecting the principle that contributions to scientific research should be evaluated independently of the person who offered the contribution. Furthermore, impossibility can be avoided by restricting the domain even without violating anonymity. Hence we do not find this route particularly attractive. We conclude that insisting on expert rights is not a promising way to avoid impossibility.

The second alternative is to avoid impossibility by allowing incomplete aggregation functions. This allows us to have a judgment aggregation function satisfying the other desiderata, at the cost of being unable to either assert or deny every proposition under consideration. A natural way to implement this suggestion is to have an aggregation function which asserts any proposition asserted by a

supermajority of the scientists in the collaboration. For instance, one could require that to assert some proposition (or its negation) at least two thirds of participating scientists must agree on the matter. Where that agreement is lacking, for instance if 51% of participating scientists are in favor of asserting the proposition and 49% against, the published unit simply remains quiet on the matter.

A judgment aggregation function that requires a sufficiently large supermajority, i.e., of $(k-1)/k$ (where k is the size of the largest minimally inconsistent subset of agenda propositions), satisfies consistency, unanimity preservation, anonymity, systematicity, acceptance/rejection neutrality and universal domain. However, among supermajority rules, the only one that satisfies deductive closure is the one that asserts only those propositions about which the individuals are unanimous (Dietrich and List 2008b, corollary 1.a). Anything short of that, and the group risks putting itself in the position of asserting in the published unit that p and that $p \rightarrow q$, but refusing to assert q.

So if we insist on deductive closure (as we have argued we should), then relaxing completeness yields an aggregation function according to which collaborating scientists can assert only those propositions about which they are unanimous (call this "unanimity rule"). Unanimity rule is thus one possible solution to the problem of judgment aggregation in science.

Moreover, unanimity rule appears to have the support of the International Committee of Medical Journal Editors (ICMJE). In a document with recommendations for authors involved in collaborative scientific projects, the ICMJE writes that "[a]ll members of the group named as authors... should have full confidence in the accuracy and integrity of the work of other group authors" (ICMJE 2013, p. 3), which we read as a requirement of unanimity. Hence there appears to be some support for unanimity preservation in scientific practice. We should note that this requirement is probably motivated by ethical rather than epistemic considerations, in particular related to assigning blame in cases of suspected fraud. The epistemic importance of fraud, however, has been highlighted in recent work in the social epistemology of science (Bruner 2013; Bright 2017). As such, there may be both an ethical and epistemic case for the ICMJE's recommendation.

Despite all this, we think that unanimity rule is too restrictive to recommend as a general rule of judgment aggregation for collaborating scientists. Based on anecdotal evidence, we think that working scientists do not regard this norm as realistic and hence ignore it. From an epistemic perspective, it seems too strict of a requirement to say that every scientist in a collaboration has to agree to a particular proposition for them to be able to assert it as a group, especially in large-scale collaborations. Thus, insofar as relaxing completeness leads us to endorsing unanimity rule as the only valid way of aggregating judgments in collaborative science, we conclude that relaxing completeness is not a particularly promising way to avoid impossibility.

This brings us to our preferred suggestion: dropping universal domain. If this is done, the other desiderata can be satisfied. In this case proposition-wise majority voting emerges as a reasonable aggregation procedure. Proposition-wise majority voting considers each proposition individually, asserting it in the published unit if the majority of individuals asserts it, and denying it if the majority denies it.

Proposition-wise majority voting satisfies unanimity preservation, anonymity, systematicity, and acceptance/rejection neutrality. It also satisfies completeness if the number of individuals is odd. When there is an even number of individuals, completeness does not hold in general as there may be ties on some propositions (but recall that we do not think completeness is normatively required).

The question is what restrictions need to be put on the domain to make sure it satisfies consistency and deductive closure as well. Dietrich and List (2010) consider a number of structural domain restrictions, the most general of which they call value restriction. Say that the individuals' judgments are value-restricted if any inconsistent set of propositions contains two propositions such that no individual asserts both of these propositions. Value restriction is sufficient for consistency (Dietrich and List 2010, proposition 7.a).

Value restriction is a sufficient condition, not a necessary one. A necessary and sufficient (but not structural) domain restriction is majority-consistency, which simply requires that there is no inconsistent subset of propositions asserted by a majority of the individuals (Dietrich and List 2010, Sect. 7). While this is not a very informative condition, it is worth emphasizing that proposition-wise majority voting may lead to a consistent published unit even when individual judgments are not value-restricted. We see no reason why a collaborative team should not be allowed to publish in such a case.

If consistency is satisfied and the number of individuals is odd, then deductive closure is also satisfied because completeness and consistency jointly imply it. If the number of individuals is even and greater than two, there are cases where proposition-wise majority voting produces a consistent but not deductively closed published unit. This can be solved by further restricting the domain. A sufficient but not necessary condition is that there are no ties. As with value restriction we do not insist on this: we simply recommend that a published unit is produced whenever proposition-wise majority voting leads to consistent and deductively closed judgments, and otherwise not.

An interesting special case occurs when there are exactly two individuals. In this case proposition-wise majority voting is identical to unanimity rule: the only way a proposition can be supported by a majority is if both individuals are willing to assert it. In this special case proposition-wise majority voting satisfies consistency and deductive closure and even universal domain, but not completeness. This seems right to us; if one is working with just one coauthor both should agree on what is written.

These results leave open the question whether some aggregation function might satisfy the same desiderata (perhaps even on a larger domain). The requirement of acceptance/rejection neutrality can be used to alleviate this worry. By Dietrich and List (2010, theorem 1), no other function than proposition-wise majority voting satisfies consistency, anonymity, and acceptance/rejection neutrality, assuming a minimally rich domain. (A minimally rich domain is one which contains all bipolar sets of individual judgments for all the propositions under consideration. A set of individual judgments is bipolar if the individuals only judge one proposition. This is a somewhat technical, but relatively weak condition.)

We conclude that, given the norms scientists already take themselves to be held to, collaborating scientists should produce papers only where they are consistent with the following judgment aggregation procedure. Use proposition-wise majority voting whenever it yields a consistent and deductively closed published unit, and do not produce a published unit at all when it does not. This satisfies all the desiderata for whose normativity we have argued.

7 Applying Our Proposal

It is interesting and somewhat surprising that the existing norms of science yield such a specific recommendation regarding the normative ideal of a judgment aggregation procedure for collaborative published units. We briefly consider two examples of how one might apply our recommendation. First, we look at one of the most direct cases of judgment aggregation procedures in science: consensus conferences. Consider this description of the procedure at a consensus conference:

> [A] group of experts called the "consensus panel"...is brought together in an open, widely advertised public forum to review and examine recent research findings. The findings are presented by the actual investigators or speakers to the conference forum in order to develop written recommendations called a consensus statement (CS) that address a number of important specific issues or consensus questions concerning a medical technology. (Wortman et al. 1988, p. 471)

Such conferences have received some attention in the philosophical literature, in particular in work by Solomon (2007, 2011, 2015) on consensus conferences in medicine. Our work here sheds new light on the puzzles raised by consensus conferences. Consensus conferences embody precisely the scenario we are modeling. Teams of researchers, clinicians, and other relevant experts come together to produce a joint document. Each person invited is supposed to have input, otherwise there would be no point inviting them. But there is no guarantee they will have precisely the same beliefs before or after the conference; if it was evident what the consensus is or should be one need not have a consensus conference. Since a consensus conference is a gathering of scientists the social norms of science should govern the behavior of participants in these conferences.

Depending on the topic of the consensus conference, non-scientists sometimes participate. We will narrow our discussion here to conferences where the aim is to settle scientific questions, rather than conferences on policy recommendations. Different norms may govern conferences with other aims.

Most obviously, our proposal suggests two things. First, consensus (read as unanimity) need not be thought of as required or preferred, as it sometimes is in practice (Solomon 2011, p. 239). Second, what is needed instead is a majority in favor of each proposition that is to end up in the consensus statement. Note also that these conferences are an occasion where it is possible to run literal votes and decide what goes into the final document based on proposition-wise majority voting, at least regarding the central questions for which the consensus conference is held. At

the same time, as emphasized above, our proposal does not require an explicit voting mechanism.

The third upshot of our proposal is a failure condition: a condition under which it would be inappropriate for the consensus conference to produce a document. As Solomon notes, it is not obvious that a brief conference of non-randomly selected experts will reach an epistemically praiseworthy consensus (Solomon 2011, p. 239). In particular, Solomon raises the worry that in our post-Kuhnian age the process of reaching consensus by means of group deliberation by scientists may be epistemically questionable, since there may not be any norms which govern how one might reach consensus that will be seen as sufficiently objective (Solomon 2007, p. 169).

An upshot of our work in Sect. 6 is that scientists should refuse to publish in cases where proposition-wise majority vote would result in a contradictory paper. If we insist that participants in consensus conferences actually hold a vote on the central questions we have a means of detecting and acting upon cases where it would be improper to issue a consensus document. This method of ruling out pseudo-consensuses is not *ad hoc*. Rather it is motivated by independent considerations regarding the social norms of science (here, in particular the norm of producing consistent published units). The possibility of such principled refusals to reach consensus can assuage worries one might have that consensus conferences generate epistemically unworthy consensuses. First, one is at least assured that any document created after a consensus conference is not a result of mere desperation. Second, since this method is motivated by pre-existing norms of science, it can share in whatever legitimacy those norms already enjoyed.

Our proposal also applies to the more everyday scientific activity of publishing coauthored papers. Our present sense of how paper writing is carried out, based on our own anecdotal experience and the observations of Wagenknecht (2015) cited in Sect. 2, is that it represents some admixture of expert rights and unanimity rule. Some participants write individual sections which are especially close to their areas of expertise, and drafts are shared with coauthors given the opportunity to veto the inclusion of propositions they are unwilling to assert (perhaps by modifying the sentences expressing them). Before we normatively appraise this, we first and foremost advocate empirical research targeted at the details of practices like these, particularly with an eye toward the judgment aggregation functions thereby implicitly embodied. It would also be worthwhile to know how and why the presently applied norms for judgment aggregation in coauthoring papers came about, since this may shed light on what challenges were addressed by the present system and hence what its epistemic advantages may be.

If it turns out that scientists occasionally or even regularly publish coauthored papers in which propositions are asserted that do not command majority support from their authors, we would advocate reform. The most direct method of ensuring that proposition-wise majority voting is followed would be to list central claims in a given draft paper, circulate that list among coauthors to solicit votes, and edit the paper in light of the majority opinion.

But this may seem somewhat unnatural to scientists, and our point is not to insist that some formalized voting procedure is implemented. As discussed in Sect. 3, we

do not mean for judgment aggregation to replace discussion or deliberation. Our focus is rather on checking that after the published unit is finished the propositions asserted therein appropriately reflect the views of the individuals involved in the collaboration (after due discussion and deliberation). We have argued that, in order to respect the pre-existing social norms of science, each proposition asserted in the published unit should command majority support from among these individuals.

8 Conclusion

We draw attention to three propositions we have argued for in this paper in order to suggest avenues for future research. First, we have argued that there is an analytic difference between collaborative teams' published units of scientific work, and the group beliefs of such teams. This distinction deserves more attention because norms appropriate when producing one may not be appropriate when producing the other. Second, we have argued that the norms working scientists accept prescribe a particular rule for aggregating judgments when producing published units: proposition-wise majority voting. Third, as a consequence of the previous two points and contrary to the advice of the ICMJE, scientists may permissibly sign their names to papers that contain propositions they do not agree with.

We believe that people can, and often do, publish statements they do not believe and may not consider themselves committed to. There exists plenty of anecdotal evidence supporting this. For example, in a recent interview, Richard Lewontin disclosed that his famous coauthored paper with the late Stephen Jay Gould, "The Spandrels of San Marco", had been mostly written by Gould alone (Wilson 2015). The Spandrels paper, as it is often called, is now recognized to be one of the most significant papers in evolutionary biology but was extremely controversial at the time of publication in 1979. In the interview, Lewontin indicated that he did not completely agree with the more polemic sections of the classic paper, which were penned by Gould exclusively.

We believe that instances like this are not uncommon in science. As scientists have an incentive to publish often, they may publish papers they believe broadly even though they disagree with their coauthors in details. While our view entails that this is sometimes permissible, note that we do not provide a blanket endorsement of this phenomenon. For example, another consequence of our proposal is that in the particular case of collaborations between two scientists, both authors should agree to everything that is written. If Lewontin's comments about the Spandrels paper are to be taken seriously, the paper should have been single-authored.

It will be instructive to see what links can be drawn between the norms for group belief and scientific publication. As noted in Sect. 2 we do not take ourselves to have directly addressed that question in this paper; we hence believe there is room for further work. If we are correct, published units may be produced the content of which is not universally assented to by all its authors. This may be thought to violate an intuitive norm of honesty. If such a norm exists it is not clear why it should be permissibly violated in science, since in single-authored papers, for instance, one

would expect the author to assent to all propositions asserted. (Note that proposition-wise majority voting also entails that in single authored papers authors should assent to all propositions, so this case cannot decide between our proposal and a putative honesty norm.) Future research may investigate whether such an honesty norm is or should be operative in science.

We note that such research on honesty norms and producing collective research documents could profitably connect up with more general prior work on the relationship between questions of epistemic responsibility, authorship, and collaborative research (Kukla 2012; Winsberg et al. 2014; Huebner et al. 2017). We have in this paper argued on the basis of egalitarian Mertonian norms that collaborating scientists should be given an equal say in final decisions as to what to assert in published research. We have not, however, touched upon questions of epistemic responsibility, implicit in any discussion of an honesty norm and made explicit in discussions of accountability for errors or task allocation in collaborative research. Whether or not our egalitarian conclusions could be sustained in light of a more complete picture of the allocation of epistemic responsibilities in collaborative research is hence an important question for evaluating the tenability of our proposal.

One striking instance of the issues raised for epistemic responsibility by our proposal concerns the role of undergraduate coauthors. If we are right, then as authors they should get equal say in deciding what propositions are asserted in the published unit. Since there may sometimes be large numbers of such researchers, they may thus be given a great degree of epistemic power, should they act collectively. If this is felt undesirable, we think this signals that the undergrads should not have been included as authors in the first place. Failing that, however, we maintain that if they did the work, their opinion should be taken seriously. Taking this on board as a practical implication of our proposal may require significant shifts in the practice of collaborative research.

We have argued that the norms scientists accept as properly governing their practice entail that, once a published unit is finished, each proposition therein should be supported by a majority of its authors. We have not argued that scientists in fact obey these norms. Future research could thus profitably investigate where scientists' practice deviates from these norms and what consequences this has for coauthored publication (see Sect. 7). We have also not argued that scientists' preferred norms are epistemically optimal. We hence also recommend future work developing metrics of epistemic optimality with which to evaluate proposition-wise majority voting against alternatives, in the context of judgment aggregation for coauthored papers.

Acknowledgements The authors thank Kevin Zollman, Christian List, Bryce Huebner, Kareem Khalifa, Thomas Boyer-Kassem, the anonymous referees, and audiences in Irvine and London for their input. LKB and RH acknowledge support from the National Science Foundation through Grant SES 1254291. RH also acknowledges support from the Leverhulme Trust and the Isaac Newton Trust through an Early Career Fellowship.

References

Anderson, M. S., Ronning, E. A., DeVries, R., & Martinson, B. C. (2010). Extending the Mertonian norms: Scientists' subscriptions to norms of research. *Journal of Higher Education*, *81*(3), 366–393. doi:10.1353/jhe.0.0095.

Annas, J. (2001). Moral knowledge as practical knowledge. *Social Philosophy and Policy*, *18*(2), 236–256. doi:10.1017/S0265052500002971. https://www.cambridge.org/core/article/moral-knowledge-as-practical-knowledge/74FBAA9E5F58E549119810BBED68368F.

Bozbay, İ., Dietrich, F., & Peters, H. (2014). Judgment aggregation in search for the truth. *Games and Economic Behavior*, *87*, 571–590, ISSN 0899-8256. doi:10.1016/j.geb.2014.02.007. http://www.sciencedirect.com/science/article/pii/S0899825614000384.

Bright, L. K. (2017). On fraud. *Philosophical Studies*, *174*(2), 291–310. doi:10.1007/s11098-016-0682-7. ISSN 1573-0883.

Bruner, J. P. (2013). Policing epistemic communities. *Episteme*, *10*(4), 403–416. doi:10.1017/epi.2013.34. http://journals.cambridge.org/article_S1742360013000348.

CDF Collaboration. (2008). Study of multi-muon events produced in p-pbar collisions at sqrt(s)=1.96 tev. Accessed online on October 3, 2016. http://arxiv.org/abs/0810.5357v1.

Cohen, L. J. (1992). *An essay on belief and acceptance*. Oxford: Clarendon Press.

Dietrich, F., & List, C. (2008a). A liberal paradox for judgment aggregation. *Social Choice and Welfare*, *31*(1), 59–78. ISSN 01761714. http://www.jstor.org/stable/41107909.

Dietrich, F., & List, C. (2008b). Judgment aggregation without full rationality. *Social Choice and Welfare*, *31*(1), 15–39. doi:10.1007/s00355-007-0260-1. ISSN 0176-1714.

Dietrich, F., & List, C. (2010). Majority voting on restricted domains. *Journal of Economic Theory*, *145*(2), 512–543. ISSN 0022-0531. doi:10.1016/j.jet.2010.01.003. http://www.sciencedirect.com/science/article/pii/S0022053110000141.

Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press.

Fagan, M. B. (2011). Is there collective scientific knowledge? Arguments from explanation. *The Philosophical Quarterly*, *61*(243), 247–269. doi:10.1111/j.1467-9213.2010.676.x. ISSN 1467-9213.

Gibbs, J. P. (1981). *Norms, deviance, and social control: Conceptual matters*. New York: Elsevier.

Gilbert, M. (1987). Modelling collective belief. *Synthese*, *73*(1), 185–204. ISSN 00397857. http://www.jstor.org/stable/20116447.

Gilbert, M. (1992). *On social facts*. Princeton: Princeton University Press.

Hartmann, S., & Sprenger, J. (2012). Judgment aggregation and the problem of tracking the truth. *Synthese*, *187*(1), 209–221. doi:10.1007/s11229-011-0031-5. ISSN 0039-7857.

Huebner, B., Kukla, R., & Winsberg, E. (2017). Making an author in radically collaborative research. In T. Boyer-Kassem, C. Mayo-Wilson, & M. Weisberg (Eds.), *Scientific collaboration and collective knowledge*. Oxford: Oxford University Press.

ICMJE. (2013). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Accessed online on October 3, 2016. http://icmje.org/icmje-recommendations.pdf.

King, C. (2013). Single-author papers: A waning share of output, but still providing the tools for progress. Accessed online on October 3, 2016. http://sciencewatch.com/articles/single-author-papers-waning-share-output-still-providing-tools-progress.

Kuhn, T. (1977). The essential tension. *The essential tension: Selected studies in scientific tradition and change* (pp. 225–239). Chicago: University of Chicago Press.

Kukla, R. (2012). "Author TBD": Radical collaboration in contemporary biomedical research. *Philosophy of Science*, *79*(5), 845–858. ISSN 00318248. http://www.jstor.org/stable/10.1086/668042.

List, C. (2005). The probability of inconsistencies in complex collective decisions. *Social Choice and Welfare*, *24*(1), 3–32. doi:10.1007/s00355-003-0253-7. ISSN 1432-217X.

List, C. (2014). Three kinds of collective attitudes. *Erkenntnis*, *79*(9), 1601–1622. doi:10.1007/s10670-014-9631-z. ISSN 1572-8420.

List, C., & Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, *18*, 89–110. ISSN 1474-0028. http://journals.cambridge.org/article_S0266267102001098.

Magnus, P. D. (2013). What scientists know is not a function of what scientists know. *Philosophy of Science*, *80*(5), 840–849. ISSN 00318248. http://www.jstor.org/stable/10.1086/673718.

Mayo-Wilson, C., Zollman, K. J. S., & Danks, D. (2011). The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, *78*(4), 653–677. ISSN 00318248. http://www.jstor.org/stable/10.1086/661777.

Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology*, *1*(1–2), 115–126.

Mulkay, M. J. (1976). Norms and ideology in science. *Social Science Information*, *15*(4–5), 637–656. 10.1177/053901847601500406. http://ssi.sagepub.com/content/15/4-5/637.short.

Priest, G. (2006). *Doubt truth to be a liar*. Oxford: Oxford University Press.

Rolin, K. (2015). Values in science: The case of scientific collaboration. *Philosophy of Science*, *82*(2), 157–177. ISSN 00318248. http://www.jstor.org/stable/10.1086/680522.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science*, *20*(1), 1–6. ISSN 00318248. http://www.jstor.org/stable/185617.

Schwitzgebel, E. (2015). Belief. In E. N. Zalta, (Ed.), *The Stanford encyclopedia of philosophy*. Summer 2015 edition. http://plato.stanford.edu/archives/sum2015/entries/belief/.

Solomon, M. (2006). Groupthink versus the wisdom of crowds: The social epistemology of deliberation and dissent. *The Southern Journal of Philosophy*, *44*(S1), 28–42. doi:10.1111/j.2041-6962.2006.tb00028.x. ISSN 2041-6962.

Solomon, M. (2007). The social epistemology of NIH consensus conferences. In H. Kincaid, & J. McKitrick (Eds.) *Establishing medical reality*, pp. 167–177. Springer Netherlands. http://dx.doi.org/10.1007/1-4020-5216-2_12.

Solomon, M. (2011). Group judgement and the medical consensus conference. In F. Gifford (Ed.), *Philosophy of Medicine* (pp. 239–254). New York: Elsevier.

Solomon, M. (2015). *Making medical knowledge*. Oxford: Oxford University Press.

Taubes, G. (1986). *Nobel dreams: Power, deceit and the ultimate experiment*. New York: Tempus Books.

Wagenknecht, S. (2015). Facing the incompleteness of epistemic trust: Managing dependence in scientific practice. *Social Epistemology*, *29*(2), 160–184. doi:10.1080/02691728.2013.794872.

Wilson, D. S. (2015). The spandrels of San Marco revisited: An interview with Richard C. Lewontin. Accessed online on October 3, 2016. https://evolution-institute.org/article/the-spandrels-of-san-marco-revisited-an-interview-with-richard-c-lewontin/.

Winsberg, E., Huebner, B., & Kukla, R. (2014). Accountability and values in radically collaborative research. *Studies in History and Philosophy of Science Part A*, *46*(1), 16–23. http://www.sciencedirect.com/science/article/pii/S0039368113001295.

Wortman, P. M., Vinokur, A., & Sechrest, L. (1988). Do consensus conferences work? A process evaluation of the NIH Consensus Development Program. *Journal of Health Politics, Policy and Law*, *13*(3), 469–498. doi:10.1215/03616878-13-3-469. http://jhppl.dukejournals.org/content/13/3/469.abstract.

Wray, K. B. (2001). Collective belief and acceptance. *Synthese*, *129*(3), 319–333. doi:10.1023/A:1013148515033. ISSN 1573-0964.

Wray, K. B. (2014). Collaborative research, deliberation, and innovation. *Episteme*, *11*:291–303. 9 ISSN 1750-0117. doi:10.1017/epi.2014.9. http://journals.cambridge.org/article_S1742360014000094.