How Can We Assess Whether to Trust Collectives of Scientists?

Elinor Clark

A great many important decisions we make in life depend on scientific information that we are not in a position to assess. So it seems we must defer to experts. By now there are a variety of criteria on offer by which non-experts can judge the trustworthiness of a scientist responsible for producing or promulgating this information. But science is, for the most part, a collective not an individual enterprise. This article explores which of the criteria for judging the trustworthiness of individual scientists can be amended for use in assessing the trustworthiness of scientific collectives. It also offers some new proposals specifically geared to assessing collective trustworthiness in science, notably, an analysis of where to apply the criteria (for example, at the group or individual level) and additional criteria to assess group infrastructure and design. The article ends with some practical suggestions for how to assist the non-expert in this task.

ORCID: orcid.org/0000-0002-5355-2963

1. Introduction

It is important that people are able to assess statements that have a direct impact on their lives. But this is challenging when these statements are about highly technical research and the individual is a non-expert without the skills, knowledge and expertise to assess the statements directly. When the public lack expertise but need to make practical decisions about a problem, many recommend they defer to experts (Meyers et al. [2020]; Meyers et al. [unpublished]) rather than trying to assess claims themselves with limited knowledge and intuition (Grundmann [2021]). Scientific research seems a prime case here: the novice is unable to understand the detailed, theory-heavy scientific research needed to assess if a statement is true, but the results matter to their lives, making it important to assess these claims (consider examples from vaccines (Goldenberg [2021]) to climate change (Almassi [2012]), as well as recent work on the challenges and importance of maintaining trust in experts in the wake of the COVID-19 pandemic (Baker [2020]; Adhikari et al. [2022])). But picking the right experts is crucial, with trust only valuable if it is directed at trustworthy agents, entities and institutions (O'Neill [2018]). So how can novices accurately identify these experts, particularly in a climate of misinformation and pseudo-experts?

A range of recent contributions help illuminate some considerations for trust in science (Wilholt [2013]; Irzik and Kurtulmus [2019]; Oreskes [2019]). One proposal is to use second-order criteria (Goldman [2001]; Anderson [2011]; Guerrero [2016]), which shifts the point of assessment from the statement to whoever is making it. This list of criteria is intended to be applied by the non-expert to assess the trustworthiness of an expert testifier.

This article focuses on applying second-order criteria to the less-explored context of collective science and how these markers may be used to assess expert groups. Recent literature explores the growth of collective science and multiple authorship, along with the wide array of practical and ethical considerations that come along with this (to name a few: Kukla [2012]; Kumar [2018]; Dang [2019]; Hosseini et al. [2022]). When considering scientific research, the relevant target for trust assessments is often not an individual scientist, but scientific collectives. This is no fringe case, with almost all scientific research conducted in groups. So, to assess the trustworthiness of this research using second-order criteria requires us to consider the application of these criteria to groups. While considerable research investigates how to use second-order criteria to assess specific experts, there is less instruction about how such assessments may be made of groups. It is this question that I address in this article.

The article proceeds as follows: Section 2 considers existing work on second-order criteria at the individual level, as well as the limits to this, and why they are important for novices making trust assessments involving statements they cannot assess directly. Section 3 considers how, in light of the collective nature of science, second-order criteria could be applied to groups, drawing on existing work from collective epistemology and group trust more generally. Section 4 raises further challenges for second-order criteria in a wider context and explores some practical recommendations for ways to ease the task of the novice, including delegation and science communication. Section 5 concludes.

2. The What, Why and How of Second-Order Criteria

2.1. Defining second-order criteria

I begin with a now familiar proposal: if non-experts cannot assess the first-order scientific statements, then we should move a level up to assess those making the statements.[1] As Anderson ([2011], p. 145) writes, 'Most laypersons cannot directly judge the merits of most scientific claims. Instead, they mostly judge what to believe by judging whom to believe'. I cannot assess the technical details describing how vaccines work, and indeed, in a well-functioning society with a good division of epistemic labour, it is not desirable that I should need to understand this. But I might be able to tell if the scientist who conducted this research seems honest, reliable and adequately qualified. Different authors reach different levels of despair about the novice's ability to make these second-order assessments. Some are more hopeful, and a number of philosophers have proposed a series of (at least in theory) accessible indicators to make second-order assessments.

Goldman ([2001], p. 93) focuses on the expert-novice problem, where a novice must identify the 'true' expert when two experts have opposing views. He highlights five sources of evidence a novice may have to make this assessment, including: dialectic abilities and the quality of experts' arguments supporting their views; agreement from additional experts; appraisals by 'meta-experts' (including appraisals reflected in formal credentials earned by the experts); evidence of the experts' interests and biases, and evidence of the experts' past track records. Since Goldman's proposal, a number of refinements and additions to the criteria have been suggested. Anderson ([2011]) proposes a second list, which she intends can be successfully applied by people with just a basic high-school education and access to the internet. These criteria include expertise, including access to relevant information and the skills and capacities to appropriately assess this information; honesty, including open communication and avoidance of selective presentation, cherry-picking data and misrepresenting opponents' views; epistemic responsibility, including following sound epistemic practices, where beliefs are formed by responsible processes and are not dogmatic, for example, meeting demands for justification by submitting to the peer-review process and noting biases and conflicts of interest and, finally, expert consensus that can be judged by assessing whether the expert has the support of a significant proportion of their peers through, for example,

[1] This can also be presented as the more general problem of how someone with no knowledge of an area is able to identify who is an expert in that area; this is a well-documented challenge, appearing under a number of names and guises as, for example, the expert-novice problem (Goldman [2001]) and the credentials problem (Cholbi [2007]; McGrath [2011]).

literature reviews, well-established institutions in the domain of expertise and surveys of trustworthy sources. Guerrero ([2016]) later distinguishes four distinct issues a novice could consider, including a scientist's expertise, their comparative expertise, their sincerity and the reliability of their testimony; and Irzik and Kurtulmus ([2019]) highlight the importance of the expert's values and moral responsibility with their concept of enhanced trust.

But the usefulness of specific criteria can be questioned. To take one example, Goldman's criteria of dialectic superiority may unfairly favour those experts who are skilled at debate and smooth-talking, while disadvantaging non-native English speakers, people with accents and groups traditionally perceived to be less authoritative. Equally, having large amounts of money to spend on science communication and creating clear, well-planned messages is likely to skew how well we perceive the expert to be communicating. Formal qualifications, which seem less interpretable, may systematically downgrade certain types of expertise that are not attained via the traditional PhD path. Track record equally faces problems. Fairly often, even highly competent scientific experts fail to achieve the results they are aiming for or fail to replicate their results (De Viterbo Pitta Borba Gouveia [2021]; Malich and Munafò [2022]): establishing a good track record can take time and what we remember and record can be highly skewed (Guerrero [2016]). Many of these criteria also require considerable inside knowledge and expertise to navigate effectively. Take honesty: spotting if a scientist is deliberately misrepresenting the views of other scientists involves knowing what those scientists are saying in the first place and catching instances of cherry-picking data would require extensive background research and insider information. Some writers argue social indicators, such as popularity, can provide valuable information (Origgi [2022]), others that the focus should solely be on a single criterion, such as consensus among scientists (Grundmann [2021]).

While an important debate, I will not try to settle which criteria are best. Instead, I will examine challenges with applying the criteria more generally. Previous work throws doubt on how applicable second-order criteria are for non-experts in practice. Not only are most people unlikely to have the time to make the intricate assessments required, some of these statements may concern beliefs important to an individual, their identity (de Cruz [2020]), and their worldview, making remaining ignorant a more desirable option (Williams [2021]). Brennan ([2020]) also emphasizes a range of internal obstacles to successful application, including cognitive biases and knowledge gaps. A number of cognitive tactics may help the novice defend against some of these biases (McGrew et al. [2019]; Wineburg et al. [2020]), though, as Watson ([2020]) points out, the intellectual humility needed to spur novices into undergoing this process of self-reflection in the first place is significant. I now turn to another question about application that has not received much attention: second-order criteria are largely discussed in relation to an individual novice assessing an individual expert, but in cases such as science, research is much more commonly produced in collectives, meaning our trust is often aimed not at single researchers, but at larger groups of scientists working together to produce the specific research.

2.2. Collective science

Large epistemically distributed collaboration is key to the success of modern scientific research (King [2012]; Huebner et al. [2014], [2017]; Dang [2019], [unpublished]). Examples abound of high-profile, large collaborations and, more generally, research collectives are the norm, often spanning numerous institutions and geographical regions. We have entered the age of thousands-large teams, such as experiments in high energy physics conducted with the Large Hadron Collider, which lists 2926 authors (The ATLAS Collaboration, Aad et al. [2008]) and widely distributed collaborations in medical research (Kukla [2012]). While research teams, for example, at university laboratories, tend to be much smaller, there is still often significant division of labour in these cases. Equally, scientists rely on each other for data, for conducting experiments and for their widely different areas of expertise, and no one individual is able to oversee and personally check the range of knowledge and data they rely on (see also Cartwright et al. [2022]).[2]

Collectives are at the heart of science. Therefore, if second-order criteria are important for trust assessments, we should be able to apply these to collectives. But the criteria suggested by Anderson and Goldman, designed as they are for cases involving individual experts, do not fully provide these tools. Are we still able to make assessments and pick out the 'true' expert when the two

[2] The picture is in fact more complex still, with technology further transforming the way scientific research is conducted, as AI systems play increasingly significant roles in the scientific discovery process (Clark and Khosrowi [2022]; Jumper et al. [2022]). Do non-expert trust assessments need to take into account the contributions of these systems when assessing the group, and what might be the implications of this shift on specific criteria and methods of making trust assessments? I will not consider these questions here, but further research on ways to make trust assessments integrating AI systems as part of the group is likely to become increasingly important as they start to play more significant roles in the research process.

'experts' with opposing views are two different research teams? Are we able to use some version of these second-order criteria to distinguish between an expert group and a pseudo-expert group? Can we judge whether to trust a piece of scientific research by assessing whether to trust the group who produced it? I turn to these questions in the next section.

First, however, note that it could be argued it is unnecessary to assess individual, or groups of, scientists specifically. Rather trust is at the level of the enterprise of science, or the scientific community as a whole, due to the robust methodologies and checks in place. I will briefly raise two challenges for this approach. First, it is not obvious that science is fully trustworthy (Contessa [2023]). For example, feminist science scholars have exposed sexism still pervasive in contemporary science (Fine [2011]; Gaston [2015]; Diogo et al. [2023];); poor incentives and competition for funding and employment for individual researchers encourage an environment where there are strong motivations for scientists to misrepresent or fabricate results (Falagas [2008]; Fanelli [2009]; Kolata [2016]); the peer-review system is far from infallible (Van Noorden [2014]; Smith [2021]) and additional biases and skews are created by private industry funding and political involvement in science (Brown [2017]; John [2020]). This all makes it unlikely that science, as currently practiced, should be taken outright to be a trustworthy institution. Second, even assuming that science as an institution is fully trustworthy, from a non-expert's perspective, it can be challenging to successfully identify which papers are 'real' science and which are fake or pseudo-science, with pseudo-science having many of the same markers and similarities in format (for example, published in seemingly reliable scientific journals, drawing on scientific methodologies). Take the case of the smoking industries' heavily funded research used to throw doubt on the link between smoking and health hazards (Christensen [2008]; Michaels [2008]), drawing on the scientific-sounding idea of 'keeping an open mind' as they called for 'more research' (Proctor [2008]). Climate sceptics also often engage with scientific arguments to support their claims (van Eck and Feindt [2022]). These cases suggest we cannot draw a neat line around science, everything inside of which novices can trust. Instead, finer-grained analysis and guidance, based on case-by-case specifics, seems needed to identify trustworthy research and research collectives.

3. Using Second-Order Criteria to Assess Groups

I argued that we often need to make trust assessments of groups when deciding whether to trust a piece of scientific research via the use of second-order assessments. So can the existing tools for second-order assessments carry across to assessments of groups? Perhaps we may hopefully think that we could just largely transfer the criteria for assessing individual scientific experts to the group context. Unfortunately, when we try to apply the foundational second-order criteria, they do not provide adequate resources for assessing groups.

One challenge is where to apply the criteria: there is no longer just one expert, but now a number of different ways to go about applying the criteria, including to each member of the group as well as the entity of the group itself. Identifying the relevant assessment point is an additional challenge for the non-expert and, if there are multiple assessments, determining how to combine these into an overall trust assessment for the group. The content of the criteria themselves also needs to be adapted to include new information specific to the collective context, such as group infrastructure and dynamics.

Turning first to where to apply second-order criteria in the group context, I will consider two broad approaches: first, applying the second-order criteria to individuals within the group and, second, applying the criteria to the group itself. I suggest that in order to gain an accurate assessment, a combination of the two approaches will be required.

3.1. Apply the second-order criteria to individuals

3.1.1. Summative approach

So to which individuals within the group should we apply the criteria? A first pass may be to say we should apply the criteria to every team member and sum up these assessments to determine if group trust is warranted. Rather like in the summative accounts of group knowledge (Quinton [1976]), we could say: Agent *x* trusts collective *y* composed of a number of individuals, *n*, if and only if *x* trusts all (or most) *n* in *y*. However, in practice, this option is not compelling.

First, applying second-order criteria to groups of scientists in this summative way is very hard work (in cases such as high energy physics and biomedical science, where the research collectives can reach thousands of individual members, simply impossible). If it was already often impractically time-consuming for a novice to use the criteria to assess a single expert, making these assessments many times over would not be workable. There is an additional challenge of determining who is in the collective in the first place. Many scientists who play important roles might not be included in author lists despite their key contributions. For example, there are a range of concerns about publication integrity with co-authored papers (Kumar [2018]), and biases, such as against female authors (West et al. [2013]), may prevent certain individuals who made important contributions from being listed. And, conversely, it can often be the case that individuals who contribute little if anything to the research end up on author lists—for example, 'big names' included to add heft and prestige to a paper (take the practice of gift, guest, and honorary authorship, which is well discussed and condemned; see Smith [1994]; Harvey [2018]). This highlights the importance of specifying the specific contributions of authors (see Holcombe [2019]; Hosseini et al. [2022]; Stewart and Hahne [2023]). And, even if we do manage to concretely define a member list of the scientific collective at the time of assessment, group membership is likely to remain in flux, with members changing and being replaced over the course of a long research project.

In addition, how well the group functions often does not boil down to a sum of how well each individual functions independently. To take one example, the group may function better if some members are excessively cautious and others prone to taking risks—while perhaps individually excessive caution, or lack of this, is a flaw, the mix of personalities can make the group as a whole work better. Another problem is that the contributions of different group members are usually unequal, but the summative approach generally treats all individuals equally. This can lead to unintuitive results. For example, it seems odd to say we do not trust the group if we trust everyone apart from a member who had little influence over the outcome, but perhaps more reasonable if the one person we do not trust is the team leader. Since the role of individuals in the group seems relevant for our trust assessments of that group, let us next consider whether only assessing 'key individuals' is a feasible alternative.

3.1.2. The 'key individuals' approach

On this approach, trust in the scientific collective equates to trust in the identified key individual(s). Although this significantly reduces the number of necessary trust assessments required, it creates a new labour-intensive and challenging task for the novice—identifying these 'key individuals'.

One plausible option is to look at the Principal Investigator (PI). So, if we trust the PI, we trust the group. The PI is a titled role, identifiable by non-experts through a (in theory) simple internet search. And they are highly influential: they play a key managerial role in designing and leading the research, ensuring its epistemic integrity and good monitoring throughout. However, the PI's role is limited: they do not have control over many aspects of the research process and are not able to assess the intentions of individual group members. And the PI can be trustworthy but the group not, and vice versa. For example, suppose the group is led by a competent, well-intentioned PI, but another important team member, one who is conducting a central part of the research, is corrupt. They convince the PI they are reliable but actually have nefarious intentions. Similarly, the PI may be individually competent and well intentioned, but the group's overarching goals or infrastructure is problematic (think, for example, of a tobacco company research team or just a group with poor communication practices and sloppy checks).

Another option for the 'key individual' could be the spokesperson. The spokesperson is a visible individual with a public-facing role that should, in theory, be accessible to non-experts. However, it can be unclear who the spokesperson is. For example, on social media, many members may be unofficial spokespeople if they are posting about the group's work, and it could be hard to distinguish comments intended in the capacity of spokesperson from those intended as personal statements. The trustworthiness of the spokesperson in honestly communicating what the group is saying also does not guarantee the trustworthiness of the collective or of the research they produce: the spokesperson may not even believe the statement they are making is true, but would still be an honest spokesperson if they accurately report the results of the group (Hawley [2017]).

Other candidates include the scientist(s) conducting the most significant or central part of the research process, the heads of the organization or each part of the project, or the person accountable for the research. But in each case, it involves a significant step to show that trusting this individual(s) adequately equates to trusting the group as a whole. Additionally, identifying these individuals likely requires information the non-expert would find difficult to access, for example untangling challenging questions about reasonable responsibility allocation. Trusting that individual (or even set of individuals), without knowing anything else about the group, does not seem to give enough information to make assessments of the group as a whole. In each case there will be examples where the 'key individual' is trustworthy, but the group is not.

Moreover, reducing group trust solely to a trust in its constituent members seems to lose something distinctive about what it means to trust a group, a point familiar from work in social epistemology (such as Tollefsen [2007]). For example, as Wilholt ([2016], p. 221) notes, Condorcet's ([1785]) jury theorem illustrates that a group can be even more competent than all of its most competent members, and no individual, or often even subset of individuals, has all the skills and information that make up the skills and competences of the whole group: the trustworthiness of a group 'crucially depends on its social organization'. As argued above, how a group interacts matters, as well as what each individual member knows. A scientific collective is not just a series of separate individuals, but a dynamic and complex interaction of different parts. Irzik and Kurtulmus ([2019]) rightly note that these considerations do not show that the trustworthiness of individual members does not impact on the group's trustworthiness. However, it does show that to make more reliable and holistic group assessments, including the group-level assessment provides extra insight we cannot get from looking at individual members alone.

To highlight this separation of group and individual, note that it can make perfect sense for someone to say that they trust group *y* but that they do not trust a number of the researchers working in *y*. This might be because while I may think these individuals have bad intentions or are not very competent or open-minded, I still trust the infrastructure of that group—for example, the checks and rules that disincentivize bad behaviour and pick up errors. And, conversely, while I may trust all the researchers in a group separately, it could still make sense not to trust the group if I think the group infrastructures are likely to outweigh or corrupt the researchers' good intentions.[3] By narrowing our assessment of trust to the individual level and failing to take into account the wider social context and organization, we fail to capture the whole picture. Equally, some criteria, such as conflicts of interests, which provide valuable and relevant information for our trust assessment, seem applicable to the group as a whole, not just the individual members.

3.2. Adapting the criteria for group application

Let us now turn to how we might make second-order assessments at the level of the group. This groups-eye-view approach has similarities to non-summative accounts of group knowledge, such

[3] Indeed, interestingly, it can be the case that the transfer works not only from trust in members to trust in groups, but also in the opposite direction—sometimes you trust a member precisely because they are part of a group.

as those proposed by Gilbert and Lackey (Gilbert [1987], [1994], [2000], [2004]; Lackey [2016]), who suggest group knowledge cannot be viewed in terms of the beliefs of its constitutive members, which are neither necessary nor sufficient for understanding the irreducibly collective beliefs of the group as a whole. Here, the focus is on the group's goals, ethos, and methodologies, rather than the individual members'.

If we are to apply the second-order criteria at the group level, the criteria being used to make the assessments of collectives will need to be adapted from the individual case. While some criteria are applicable to groups in the same way as to individuals, for example, whether the group's research is in line with expert consensus, its informal reputation and whether it follows sound epistemic practices, submits to the peer-review process and makes itself accountable, other criteria work less well in the group context. Expertise and qualifications pose a challenge: the group as a whole has not been awarded a PhD or other academic qualification. Track record is also limited when applied at the group level: looking only at research that a group produces together, and not what has been produced by group members outside the collective, misses important information and unnecessarily restricts our assessments. This challenge is exacerbated as many groups conducting scientific research are not well-established collectives. They are often in flux and can spring up spontaneously without specific ethos and intent. For example, a newly formed group, or a group that exists solely for the duration of a specific project, will have no track record or reputation, and it may be challenging for a novice to assess whether it is following markers of epistemic responsibility and honesty.

There are also other useful criteria to consider at the group level, including institutional infrastructure and the dynamics of the group, as others have suggested (for example, Gundersen and Holst [2022]). Using the conditions suggested by Gunderson and Holst, along with the relevant original criteria, as an initial base, I will now briefly outline some possibilities for group-specific second-order criteria.[4]

(1) Institutional independence and transparency of group infrastructure. Does the group state any conflicts of interest (which are relevant at the group, as well as individual, level)? Are they open about internal deliberation and decision-making processes (Elliott [2022]), when

[4] I thank an anonymous reviewer for pushing me to consider more explicitly specific criteria for group assessments.

this is appropriate?[5] Is the group under undue influence from political or commercial interests?

(2) Reliable epistemic practices and group dynamics. Does the group follow good epistemic practices, openly submitting to peer review and avoiding dogmatic thinking? Are there good group dynamics; for example, is there one more powerful member who exerts excessive control over the group, with objections of others consistently ignored? Are there incentives within the group that encourage sloppy research?[6]

(3) As with the individual, indicators such as whether the research produced by the group is in line with scientific consensus, is communicated openly and honestly and has a good track record (if this exists) are also important. For example, looking at the track record of the group can provide relevant information about how the team works together beyond the track record of the individual scientists.

More research fleshing out these group-specific criteria and how they may be made accessible to non-experts is required. But for now, let us turn to assessing how a novice could go about applying such group-oriented second-order criteria to scientific collectives in practice. Here, a number of additional challenges arise.

3.3. Challenges with group-level assessments

One first challenge, as gestured to above, is how to identify the specific group to assess. This is especially complex when considering highly diffuse collaborations spread out across numerous teams with varying working methods, conditions and intentions and limited cross-team interaction. Many scientists working on large, epistemically distributed projects are working on widely different and non-overlapping questions. Indeed, there can often be significant disagreement within

---

5 While transparency seems an important marker to factor into trust assessments of groups, there may be cases where lack of transparency needn't be a reason not to trust a group and its findings (for example, if the findings contain particularly sensitive information).

6 Group dynamics could also be assessed through statements about group 'culture', which can in some cases act as a proxy to understand how the group, at least tries, to operate (though of course this should not necessarily be taken on face value).

groups (Dang [unpublished]). So how can we assess the group 'as a whole' if it is made up of such a diverse and often conflicting group of individuals?

There is also a challenge, particularly within large, hierarchical organizations that have teams within teams, in deciding at which level to make the assessment. Take, for example, research produced by a large corporation, such as Google's DeepMind. How fine-grained should the assessments of this group be: is the relevant collective the specific department that research project is conducted in or the smaller team of key researchers working on it? For example, two teams of researchers working in the same departments may be trusted to different extents, requiring a more fine-grained assessment at the level of the group itself. Both the narrower team-level assessments, and the wider social and institutional context in which they are operating, seem relevant for the overall trust assessment.

Another challenge is that, just as the group as a whole can make a difference to trust assessments beyond its individual members, individual members can also make a difference beyond the group. Say a scientific research collective was established with virtuous goals and has previously been epistemically responsible, producing high-quality research for the public good. But you happen to know that the research director has recently been replaced by a scientist with dubious credentials, who you have reason to believe has conflicting interests. Knowing this, it might not make sense to continue to trust the group, even if its stated goals are still honourable and its track record still exemplary. It seems important to also take into account whether individuals in the collective are trustworthy, meaning the analysis at the individual level may still be important to determine whether trust is warranted. Additionally, as mentioned above, certain crucial criteria, such as those judging scientific expertise, are hard to apply at the level of the group, meaning the individual level assessments are still important.

I have argued that assessing solely at the level of individuals or solely at the group level is unlikely to lead to an accurate and full assessment. Perhaps this is not surprising. Instead, a combined approach seems the most promising. But how are we to know how many assessments are required and how to weigh each separate assessment in the overall trust rating? Should each component weigh equally or should some count for more than others? I suspect the answer is likely to be conditional on the specific case. For example, some smaller groups with limited track record, which are largely led by one influential scientist, would require that scientist's assessment to feed heavily into the final assessment, whereas for groups with a clearer overarching group structure, such as well-established bodies like the IPCC,[7] the group-level assessment may be the biggest part of the overall assessment.

As we have seen, making assessments of groups involves not just more work (for example, applying the criteria numerous times) but also different types of work and complexity for the novice. Applying the criteria to groups rather than an individual involves more points of assessments. It also brings in new criteria to capture the trustworthiness of the collective and the task of identifying the group to which to apply these criteria. Identifying the key individuals and how many individuals to assess is another task. Finally, once these assessments have been roughly made, how are the assessments to be weighted to make a final trust judgement? A general prescription about how to combine these assessments is unlikely to work, with the appropriate weightings likely to vary case-by-case.

A total fulfilment of the criteria is not generally necessary for trust (Gunderson and Holst [2022]), and some criteria may carry more weight than others. Some criteria seem consistently important, such as having relevant and competent scientific expertise. But, in some cases, the usefulness of indicators such as transparency may be restricted to activities like submitting to peer review, particularly if the research is potentially harmful to the public or proprietary. Equally, there may be cases where, despite imbalanced group dynamics and lack of diversity, trust in the collective is still warranted, even if not ideal.

I have been working in this article at a rather ideal level and, as has been pointed out with critiques of second-order criteria previously, the information needed for these criteria is often unrealistic for a novice to be able to access. It is also worth pointing out that, as Gunderson and Holst ([2022]) show, trust and trustworthiness often do not match up and institutions can be highly trusted without being trustworthy according to the suggested criteria. The reality is messier still and arguing that the public should use the criteria clearly does not mean that they will do (to take one in-practice example, see Eysenbach and Kahler [2002]). I will end by considering some final wider challenges to consider for further research, along with some ways it may be possible to ease the assessment work in practice.

---

[7] Though note the IPPC is not a scientific collective per se, due to its more political and advisory role, and its focus on assessing scientific literature, rather than conducting its own scientific research.

4. Further Research and a Helping Hand

4.1. Delegate the work

As we can see from the preceding discussion, there is a significant amount of work needed to make these combined assessments, a large amount of insider information to access, as well as significant intuition required to judge how to weigh the assessments and how many are needed. This is a challenging task that may be more suited to trained, professional 'expert assessors' who have easier access to the necessary information, have dedicated time to make the assessments carefully and are trained with the skills and knowledge to apply the criteria effectively. The details of these assessments could then be made available to the public so they can read into the findings themselves. Delegation is often essential (Nguyen [2018]; Lillehammer [2021]) in many areas of life, due to lack of time, knowledge and resources, and the specificity of expertise, and could be an effective aid in this case.

Effectively setting up a system delegating to expert assessors brings a range of questions: how can we monitor the quality of an expert checker's assessments or, in other words, do we need expert checkers to check the expert checkers?[8] How can we ensure the information found by the expert assessors and their application of the second-order criteria is displayed publicly to allow transparency and easy searchability for non-experts? Are there ways the expert assessors could publicly display their own credentials, track record and expertise? How can we ensure the correct incentives are in place to prevent expert assessors choosing to mislead or exaggerate, rather like a science journalist selling a story?[9] How can we ensure a diverse group of expert assessors? Will any of this enable us to avoid the principal-agent problem and ensure the expert assessors are (at least largely) acting in our epistemic interests? Inevitably, many groups will not trust the assessments of these expert assessors. But it may provide a way of enabling easier access of information

[8] It is worth noting that these 'expert assessors' may not only be individuals but also groups, such as work centres and think tanks, that assess science in order to make recommendations for a range of policy questions. I will not look into this further here, but a full discussion would include their contributions and would need to be acknowledged.

[9] It should also be noted that relying on the scientists to communicate directly can also be a highly biased communication method, with scientists seeking to promote their own agendas and to present their research in the best light, led by the strong incentives of funding and career progression. See episode 3, 'Thank You for Not Smoking, of *The Trust Race* podcast (available at <open.spotify.com/episode/3KArKq96fXJM0eNjnIxjzo>).

and assistance in making assessments that otherwise are likely to be an excessive burden on non-experts.

Another way delegation could take place is to run randomly selected novice citizen panels (Crosby et al. [1986]; Norheim et al. [2021]). These panels could make assessments using some version of the second-order criteria, with other non-experts then able to input this assessment into their own decisions about whom to trust. However, it is unclear how useful these panels are. For example, some damaging critiques of the effectiveness of citizen panels (for example, Kathlene and Martin [1991]) throw light on how participants can be easily led, vulnerable to biases and bad design processes, which ultimately skews outcomes. Panel meetings are also labour intensive and time consuming for those participating, making them impractical for the majority of cases. But it may be that well-designed citizens' panels could provide valuable input into the process of assessing which scientists to trust. Both these forms of delegation also assume that the problems raised above with assessing groups can be (at least partly) overcome with more time, help and expertise thrown at the challenge.

4.2. Science communication

But even if we assume there are neutral expert checkers acting solely to provide accurate information independent of personal agendas and other external influences, will non-experts access their information directly from these individuals or will it be mediated through other outlets that may be biased or sloppy (for example, Contessa [2022])? In particular, science journalists are often this extra link. So yet another trust assessment needs to be made further down the line. We can invest great time and care in researching the credentials, honesty and epistemic integrity of a group and its infrastructures, but if this research is then distorted and sensationalized when it is communicated to the public, then the information that ends up being consumed will not be trustworthy. This pushes consideration of how to assess different news outlets as a very important point of assessment (perhaps joining up with expert assessors, as discussed above), as noted by scholars like Anderson ([2011]).

Other ways to assist the application of second-order criteria in practice is by improving how information is presented for the public to assess. For example, there could be easy-to-access, public listings of key group members to aid identification or amalgamated listings about funding, conflicts of interest and peer-review outcomes by a publicly funded body and collected information relating to specific criteria, such as consensus, which could even be presented in easy-to-access and already well-trusted sites, such as Wikipedia (Okoli et al. [unpublished]).

4.3. Consensus and many-analysts research

Building on this, consensus seems a particularly important criteria and communicating a scientific consensus can be influential for building public trust on a range of issues (Bialek et al. [2023]).[10] However, identifying consensus is also challenging for non-experts and can be misconstrued due to the visibility of 'merchants of doubt' (Oreskes and Conway [2010]). A number of new proposals may help the public more easily get a hold of scientific consensus on a range of questions, such as the fledgling institute to measure or establish scientific consensus (IASC) (Vickers [2022], [forthcoming]; Adam [2023]). If this information were available for non-experts to search in an accessible format, this could ease the challenging task of collecting this information on a case-by-case basis.

There are additional reasons for considering consensus as more reliable than single assessments of individual teams. In particular, considering recent work on the many-analysts cases suggests that even expanding to the group level is often not wide enough to capture a reliable assessment and there are cases where well-qualified groups all come to different conclusions, even starting from the same dataset.[11] For example, in (Silberzahn et al. [2018]), twenty-nine teams of analysts were given the same dataset to answer the research question: are soccer referees more likely to give red cards to dark-skin-toned players than to light-skin-toned players? The approaches different teams took and the results found varied widely, from finding a statistically significantly positive effect to observing no significant relationship. Different findings may result from a range of reasons including differences in choice of primary data, choice of averaging technique and choice of effect measure, and will also be influenced by personal and financial interests (Stegenga [2011]). These findings are disturbing: we are often encouraged in our current system of publishing and

[10] Though note communicating consensus can also come with its own range of challenges and may not always promote trust, perhaps giving the impression that the question has been poorly or narrowly investigated or that there is some political agenda.

[11] I thank Peter Vickers and Haixin Dang for bringing this point to my attention.

communicating science to focus on single results and strong narratives (Silberzahn and Uhlmann [2015]). This also problematically skews to promoting 'interesting results', leading to publication biases for statistically significant positive findings over null or negative findings (Brown [2008]). These findings from many-analysts cases suggest that significant variation in findings when analysing complex data may be difficult to avoid, even by well-qualified experts with honest intentions. One upshot is that even if expert groups come out as trustworthy according to the second-order criteria assessments, this doesn't mean we should trust their findings. Instead, the more reliable method may be to adopt some kind of crowd-sourcing of analysis, where a number of research teams simultaneously investigate a research question, and wait until several teams have looked at the issue to see if a consensus of opinion forms. Crowd-sourcing may not be the ideal solution in all cases, requiring large amounts of time and effort. However, particularly in cases where the research has direct importance, for example to policy questions, crowd-sourcing may be a valuable approach, allowing analysis choices to be explored, reducing incentives for 'interesting' over accurate results and providing discussion space for dissent (Silberzahn and Uhlmann [2015]).

Just as it seems important to situate individual scientists within their group context, so we should consider that the non-experts making these assessments are not operating alone but within wider epistemic environments with a range of possibilities for connection and assistance. More research is needed from across disciplines including policy, philosophy, education, science and communication studies, to simplify and ease the task of applying second-order criteria to groups.

5. Conclusion

In this article I have highlighted that when trying to assess scientific research as non-experts, we may often want to be able to apply some kind of second-order criteria to collectives. However, I have shown that this application to groups brings a range of new challenges in practice, including increasing the number of assessments and necessitating new criteria and types of assessment. I have considered current work on second-order criteria and group trust assessments and outlined a set of potential criteria for assessing groups. I investigated ways of applying second-order criteria to scientific collectives in practice and suggested a combined approach, using assessments of both individual members and the group itself, is likely to be needed to make trust assessments of groups.

I ended by adding some extra considerations and practical ways forward for the non-expert, including looking at delegation, science communication and consensus. Developing accounts that better integrate a range of trust assessments more easily and acknowledging that assessments of a piece of scientific research using second-order criteria may involve assessments of a range of actors, entities and institutions including individual scientists and scientific collectives, is likely to only become increasingly important.

Acknowledgements

Thank you to two anonymous reviewers for the *British Journal for the Philosophy of Science*, Peter Vickers, Nancy Cartwritght, Sam Colclough, Deb Marber, members of CHESS working group, attendees of PERITA 2023 conference, James Bell, Tanja Rechnitzer, Torsten Wilholt: I greatly appreciate your insightful comments at different stages of writing of this article. I am also grateful for doctoral funding from the Arts and Humanities Research Council-funded Northern Bridge Consortium.

Department of Philosophy
Durham University
Durham, United Kingdom
elinor.s.clark@durham.ac.uk

References

Adam D. [2023]: ‘Can Giant Surveys of Scientists Fight Misinformation on COVID, Climate Change, and More?’, *Nature*, **617**, pp. 452–54.

Adhikari B., Yeong Cheah P. and von Seidlein L. [2022]: ‘Trust Is the Common Denominator for COVID-19 Vaccine Acceptance: A Literature Review’, *Vaccine: X*, **12**, available at <doi.org/10.1016/j.jvacx.2022.100213>.

Almassi, B. [2012]: ‘Climate Change, Epistemic Trust, and Expert Trustworthiness’, *Ethics and the Environment*, **17**, pp. 29–49.

Anderson, E. [2011]: ‘Democracy, Public Policy, and Lay Assessments of Scientific Testimony’, *Episteme*, **8**, pp. 144–64.

The ATLAS Collaboration, Aad, G., Abat, E., Abdallah, J., Abdelalim, A. A., Abdesselam, A., Abdinov, O. et al. [2008]: ‘The ATLAS Experiment at the CERN Large Hadron Collider’, *Journal of Instrumentation*, **3**, available at <doi.org/10.1088/1748-0221/3/08/S08003>.

Baker D. W. [2020]: 'Trust in Health Care in the Time of COVID-19', *Journal of the American Medical Association*, **324**, pp. 2373–75.

Białek, M., Meyers, E. A., Arriaga, P., Harateh, D., Urbanek, A. [2023]: 'COVID-19 Vaccine Skeptics Are Persuaded by Pro-vaccine Expert Consensus Messaging', *Journal of Experimental Psychology: Applied*, **29**, pp. 477–88.

Brennan, J. [2020]: 'Finding the Snark Together: A Response to Watson and Hinton', *Social Epistemology Review and Reply Collective*, **9**, pp. 54–59.

Brown, J. [2008]: 'The Community of Science', in M. Carrier, D. Howard, and J. A. Kourany (*eds*), *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*, Pittsburgh, PA: University of Pittsburgh Press.

Brown, J. [2017]: 'Socializing Medical Research', in K. C. Elliott and D. Steel (*eds*), *Current Controversies in Values and Science*, Abingdon: Routledge, pp. 147–60.

Cartwright, N., Hardie, J., Montuschi, E., Soleiman, M. and Thresher, A. [2022]: *The Tangle of Science: Reliability beyond Method, Rigour, and Objectivity*, Oxford: Oxford University Press.

Christensen, J. [2008]: 'Smoking out Objectivity: Journalistic Gears in the Agnotology Machine', in R. Proctor and L. L. Schiebinger (*eds*), *Agnotology: The Making and Unmaking of Ignorance*, Stanford University Press, pp. 266–82.

Cholbi, M. [2007]: 'Moral Expertise and the Credentials Problem', *Ethical Theory and Moral Practice*, **10**, pp. 323–34.

Clark, E. and Khosrowi, D. [2022]: 'Decentring the Discoverer: How AI Helps Us Rethink Scientific Discovery', *Synthese*, **200**, 463, available at <doi.org/10.1007/s11229-022-03902-9>.

Condorcet, M. [1785]: Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix (PNG) (in French)

Contessa, G. [2022]: 'Shopping for Experts', *Synthese*, **200**, available at <doi.org/10.1007/s11229-022-03590-5>.

Contessa, G. [2023]: 'It Takes a Village to Trust Science: Towards a (Thoroughly) Social Approach to Public Trust in Science', *Erkenntnis*, **88**, pp. 2941–66.

Crosby, N., Kelly, J. M. and Schaefer, P. [1986]: 'Citizens Panels: A New Approach to Citizen Participation', *Public Administration Review*, **46**, pp. 170–78.

Dang, H. [2018]: 'Epistemic Responsibility in Science', available at <www.haixindang.com/uploads/5/9/8/4/59847021/dang_new_handout.pdf>, accessed 2 April 2022.

Dang, H. [2019]: 'Epistemology of Scientific Collaborations', PhD Thesis, University of Pittsburgh.

De Cruz, H. [2020]: 'Believing to Belong: Addressing the Novice–Expert Problem in Polarized Scientific Communication', *Social Epistemology*, **34**, pp. 440–52.

De Viterbo Pitta Borba Gouveia, S. [2021]: 'The Limits of Randomised Controlled Trials: Lessons from the Replication Crisis', *Rerum Causae*, **12**, available at <rc.lse.ac.uk/articles/182>.

Diogo, R, Adesomo, A, Farmer, K. S., Kim, R. J., Jackson, F. [2023]: 'Not Just in the Past: Racist and Sexist Biases Still Permeate Biology, Anthropology, Medicine, and Education', *Evolutionary Anthropology*, **32**, pp. 67–82.

Elliott, K. [2022]: 'A Taxonomy of Transparency in Science', *Canadian Journal of Philosophy*, **52**, pp. 342–55.

Eysenbach, G. and Köhler, C. [2002]: 'How Do Consumers Search for and Appraise Health Information on the World Wide Web? Qualitative Study Using Focus Groups, Usability Tests, and In-Depth Interviews', *British Medical Journal*, **324**, available at <doi.org/10.1136/bmj.324.7337.573>.

Falagas, M. E. and Alexiou, V. G. [2008]: 'The Top-Ten in Journal Impact Factor Manipulation', *Archivum Immunologiae et Therapiae* Experimentalis, **56**, pp. 223–26.

Fanelli, D. [2009]: 'How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-analysis of Survey Data', *PLOS One*, **4**, available at <doi.org/10.1371/journal.pone.0005738>.

Fine, C. [2011]: *Delusions of Gender: How Our Minds Society and Neurosexism Create Difference*, New York: W. W. Norton.

Gaston, N. [2015]: *Why Science Is Sexist*, Wellington: Bridget Williams.

Gilbert, M. [1987]: 'Modelling Collective Belief', *Synthese*, **73**, pp. 185–204.

Gilbert, M. [1994]: 'Remarks on Collective Belief', in F. F. Schmitt (*ed.*), *Socializing Epistemology: The Social Dimensions of Knowledge*, Lanham, MD: Rowman and Littlefield, pp. 235–56.

Gilbert, M. [2000]: *Sociality and Responsibility: New Essays in Plural Subject Theory*, Lanham, MD: Rowman and Littlefield

Gilbert, M. [2004]: 'Collective Epistemology', *Episteme*, **1**, pp. 95–107.

Goldenberg, M. J. [2021]: *Vaccine Hesitancy: Public Trust, Expertise, and the War on Science*, Pittsburgh, PA: University of Pittsburgh Press.

Goldman, A. I. [2001]: 'Experts: Which Ones Should You Trust?', *Philosophy and Phenomenological Research*, **63**, pp. 85–110.

Goldman, A. I. [2018]: 'Expertise', *Topoi*, **37**, pp. 3–10.

Grundmann, T. [2021]: 'Experts: What Are They and How Can Laypeople Identify Them?', in J. Lackey and A. McGlynn (*eds*), *Oxford Handbook of Social Epistemology*. Oxford University Press.

Guerrero, A. A. [2016]: 'Living with Ignorance in a World of Experts', in R. Peels (*ed.*), *Perspectives on Ignorance from Moral and Social Philosophy*, New York: Routledge, pp. 156–85.

Gundersen, T. and Holst, C. [2022]: 'Science Advice in an Environment of Trust: Trusted, but Not Trustworthy?', *Social Epistemology*, **36**, pp. 629–40.

Harvey, L. [2018]: 'Gift, Honorary, or Guest Authorship', *Spinal Cord*, **56**, available at <doi.org/10.1038/s41393-017-0057-8>.

Hawley, K. [2017]: 'Trustworthy Groups and Organizations', in P. Faulkner and T. Simpson (*eds*), *The Philosophy of Trust*, Oxford: Oxford University Press, pp. 230–50.

Holcombe, A. O. [2019]: ‘Contributorship, Not Authorship: Use CRediT to Indicate Who Did What’, *Publications*, **7**, <doi.org/10.3390/publications7030048>.

Hosseini, M., Lewis, J., Zwart, H. and Gordijn, B. [2022]: ‘An Ethical Exploration of Increased Average Number of Authors per Publication’, *Science and Engineering Ethics*, **28**, <doi.org/10.1007/s11948-021-00352-3>.

Huebner, B., Kukla, R. and Winsberg, E. [2014]: ‘Accountability and Values in Radically Collaborative Research’, *Studies in the History and Philosophy of Science Part A*, **46**, pp. 16–23.

Huebner, B., Kukla, R. and Winsberg, E. [2017]: ‘Making an Author in Radically Collaborative Research’, in T. Boyer-Kassem, C. Mayo-Wilson and M. Weisberg (*eds*), *Scientific Collaboration and Collective Knowledge*, Oxford: Oxford University Press, pp. 95–116.

Irzik, G. and Kurtulmus, F. [2019]: ‘What Is Epistemic Public Trust in Science?’, *British Journal for the Philosophy of Science*, **70**, pp. 1145–66.

John, S. [2020]: ‘Should Science Lead?’, *The Philosophers’ Magazine*, **90**, pp. 58–63.

Jumper, J., Evans, R., Pritzel, A. et al. [2021]: ‘Highly Accurate Protein Structure Prediction with AlphaFold’, *Nature*, **596**, pp. 583–89.

Kathlene, L. and Martin, J. A. [1991]: ‘Enhancing Citizen Participation: Panel Designs, Perspectives, and Policy Formation’, *Journal of Policy Analysis and Management*, **10**, pp. 46–63.

King, C. [2012]: ‘Multiauthor Papers: Onward and Upward’, *ScienceWatch Newsletter*, July 2012, available at <archive.sciencewatch.com/newsletter/2012/201207/multiauthor_papers/>.

Kolata, G. [2016]: ‘So Many Research Scientists, So Few Openings as Professors’, *The New York Times*, 14 July 2016, available at <www.nytimes.com/2016/07/14/upshot/so-many-research-scientists-so-few-openings-as-professors>.

Kukla, R. [2012]: ‘“Author TBD”: Radical Collaboration in Contemporary Biomedical Research’, *Philosophy of Science*, **79**, pp. 845–58.

Kumar, S. [2018]: ‘Ethical Concerns in the Rise of Co-authorship and Its Role as a Proxy of Research Collaborations’, *Publications*, **6**, available at <doi.org/10.3390/publications6030037>.

Lackey, J. [2016]: ‘What Is Justified Group Belief’, *Philosophical Review Recent Issues*, **125**, pp. 341–96.

Lillehammer, H. [2021]: ‘Testimony, Deference, and Value’, in M. Hannon and J. de Ridder (*eds*), *The Routledge Handbook of Political Epistemology*, London: Routledge. pp. 458–68.

Malich, L. and Munafò, M. R. [2022]: ‘Replication of Crises: Interdisciplinary Reflections on the Phenomenon of the Replication Crisis in Psychology’, *Review of General Psychology*, **26**, available at <doi.org/10.1177/10892680221077997>.

McGrath, S. [2011]: ‘Skepticism about Moral Expertise as a Puzzle for Moral Realism’, *The Journal of Philosophy*, **108**, pp. 111–37.

McGrew, S., Smith, M., Breakstone, J., Ortega, T. and Wineburg, S. [2019]: ‘Improving University Students’ Web Savvy: An Intervention Study’, *British Journal of Educational Psychology*, **89**, pp. 485–500.

Meyers, E. A., Gretton, J. D., Budge, J., Fugelsang, J. and Koehler, D. J. [unpublished]: 'Broad Effects of Shallow Understanding: Explaining an Unrelated Phenomenon Exposes the Illusion of Explanatory Depth', available at <doi.org/10.31234/osf.io/3krx8>.

Meyers, E. A., Turpin, M. H., Białek, M., Koehler, D. J. and Fugelsang, J. A. [2020]: 'Inducing Feelings of Ignorance Makes People More Receptive to Expert (Economist) Opinion', *Judgment and Decision Making*, **15**, pp. 909–25.

Michaels, D. [2008]: 'Manufactured Uncertainty: Contested Science, and the Protection of the Public's Health and Environment', in R. N. Proctor and L. Schiebinger (*eds*), *Agnotology: The Making and Unmaking of Ignorance*, Stanford, CA: Stanford University Press, pp. 90–107.

Nguyen, C. T. [2018]: 'Expertise and the Fragmentation of Intellectual Autonomy', *Philosophical Inquiries*, **6**, pp. 107–24.

Norheim, O. F., Abi-Rached, J. M., Bright, L. K. et al. [2021]: 'Difficult Trade-Offs in Response to COVID-19: The Case for Open and Inclusive Decision Making', *Nature Medicine*, **27**, pp. 10–13.

Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F. Å and Lanamäki, A. [unpublished]: 'The People's Encyclopedia under the Gaze of the Sages: A Systematic Review of Scholarly Research on Wikipedia', available at <dx.doi.org/10.2139/ssrn.2021326>.

O'Neill, O. [2018]: 'Linking Trust to Trustworthiness', *International Journal of Philosophical Studies*, **26**, pp. 293–300.

Oreskes, N. [2019]: *Why Trust Science?* Princeton, NJ: Princeton University Press.

Oreskes, N. and Conway, E. M. [2010]: *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*, New York: Bloomsbury.

Origgi, G [2022]: 'The Social Indicators of the Reputation of an Expert', *Social Epistemology*, **36**, pp. 541–49.

Proctor, R. [2008]: 'Agnotology: A Missing Term to Describe the Cultural Production of Ignorance (and Its Study)', in R. Proctor and L. L. Schiebinger (*eds*), *Agnotology: The Making and Unmaking of Ignorance*, Stanford, CA: Stanford University Press, pp. 1–35.

Quinton, A. [1976]: 'Social Objects', *Proceedings of the Aristotelian Society*, **75**, pp. 1–27.

Silberzahn, R. and Uhlmann, E. [2015]: 'Crowdsourced Research: Many Hands Make Tight Work', *Nature*, **526**, pp. 189–91.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š. et al. [2018]: 'Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results', *Advances in Methods and Practices in Psychological Science*, **1**, pp. 337–56.

Smith, E. M. [2021]: 'Reimagining the Peer-Review System for Translational Health Science Journals', *Clinical and Translational Science*, **14**, pp. 1210–21.

Smith, J. [1994]: 'Gift Authorship: A Poisoned Chalice?', *British Medical Journal*, **309**, pp. 1456–7.

Stegenga, J. [2011]: 'Is Meta-analysis the Platinum Standard of Evidence?', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **42**, pp. 497–507.

Stewart Jr, C. N. and Hahne, G. [2023]: ‘Authorship and the Importance of the Author Contribution Statement’, *Plant Cell Reports*, **42**, pp. 655–56.

Tollefsen, D. [2007]: ‘Group Testimony’, *Social Epistemology*, **21**, pp. 299–311.

van Eck, C. W. and Feindt, P. H. [2022]: ‘Parallel Routes from Copenhagen to Paris: Climate Discourse in Climate Sceptic and Climate Activist Blogs’, *Journal of Environmental Policy and Planning*, **24**, pp. 194–209.

Van Noorden, R. [2014]: ‘Publishers Withdraw More Than 120 Gibberish Papers’, *Nature*, available at <doi.org/10.1038/nature.2014.14763>.

Vickers, P. [2022]: *Identifying Future-Proof Science*, Oxford: Oxford University Press.

Vickers, P. [forthcoming]: ‘Don’t We All Believe in Scientific Facts? Replies to My Critics’, *Metascience*, available at <doi.org/10.1007/s11016-023-00926-w>.

Watson, J. C. [2020]: ‘Hunting the Expert: The Precarious Epistemic Position of a Novice’, *Social Epistemology Review and Reply Collective*, **9**, pp. 51–58.

West, J. D., Jacquet, J., King, M. M., Correll, S. J. and Bergstrom, C. T. [2013]: ‘The Role of Gender in Scholarly Authorship’, *PLOS One*, **8**, available at <doi.org/10.1371/journal.pone.0066212>.

Wilholt, T. [2013]: ‘Epistemic Trust in Science’, *British Journal for the Philosophy of Science*, **64**, pp. 233–53.

Wilholt, T. [2016]: ‘Collaborative Research, Scientific Communities, and the Social Diffusion of Trustworthiness’, in M. S. Brady and M. Fricker (*eds*), *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*, Oxford: Oxford University Press, pp. 218–34.

Williams, D. [2021]: ‘Motivated Ignorance, Rationality, and Democratic Politics’, *Synthese*, **198**, pp. 7807–27.

Wineburg, S., Breakstone, B., Ziv, N. and Smith, M. [2020]: ‘Educating for Misunderstanding: How Approaches to Teaching Digital Literacy Make Students Susceptible to Scammers, Rogues, Bad Actors, and Hate Mongers’, Working Paper, A-21322, Stanford History Education Group, available at <purl.stanford.edu/mf412bt5333>.