

# 11

## Peer Review

PAUL N. EDWARDS

### Overview

Despite many flaws, including variable quality and a lack of universal standards, peer review – the formal process of critically assessing knowledge claims prior to publication – remains a bedrock norm of science. It therefore also underlies the scientific authority of the Intergovernmental Panel on Climate (IPCC). Most literature used in IPCC assessments has already been peer reviewed by scientific journals. IPCC assessments are themselves reviewed at multiple stages of composition, first by Lead Authors (LAs), then by scientific experts and non-governmental organisations outside the IPCC, and finally by government representatives. Over time, assessment review has become increasingly inclusive and transparent: anyone who claims expertise may participate in review, and all comments and responses are published after the assessment cycle concludes. IPCC authors are required to respond to all comments. The IPCC review process is the most extensive, open and inclusive in the history of science. Challenges include how to manage a huge and ever-increasing number of review comments, and how to deal responsibly with review comments that dispute the fundamental framing of major issues.

### 11.1 Introduction

The IPCC's claim to scientific authority is heavily based on the multiple levels of peer review applied in its assessments. Peer review practices date to the 1730s, if not even earlier (Spier, 2002). They are a deeply entrenched norm, based in the fundamental scientific principles of communal knowledge production and methodological scepticism. When the IPCC was established in 1988, journal review systems had acquired a stable form, which remains prevalent today. Scientists submit articles to journals. Journal editors then locate referees, who write

reviews detailing errors, methodological issues or other problems and recommend either rejection, revision or acceptance. The most common recommendation is to revise and resubmit. If the referees and editor agree that the revisions respond adequately to their comments, the paper is accepted.

Scholars have studied peer review systems for decades. Studies have unearthed problems ranging from failure to catch obvious mistakes to favouritism ('pal review') to outright fraud (Chubin & Hackett, 1990; Moran, 1998). It's a messy and imperfect process – and in practice there are few, if any, universal standards. Both referees and editors face time and expertise constraints, which lead to widely varying levels of investment in the process. Different journals require double-blind (neither referees nor authors know each other's names), single-blind (referees know the authors' names, but not vice versa), signed or optionally signed reviews. Many journals require referees to answer specific questions or fill out rating scales, but these are weak checks on an inherently qualitative process. In practice, reviews run the gamut from brief, pro forma recommendations to multi-page deep dives into methods, mathematics and supporting or conflicting literature.

A key weakness: unlike auditors in banking and corporate finance, peer reviewers rarely attempt to replicate or test any part of a study (McIntyre & McKittrick, 2005). They rely instead on their expert knowledge, and they assume the good faith and honesty of authors. This honour system has led to scandals in some fields when formal replication studies have disconfirmed results previously held as fundamental (Baker, 2015).

## 11.2 Who Counts as a Peer?

In journal review systems, 'peers' are generally understood to be experts in the same or closely related fields. Editors' choice of peers can influence publication decisions. Yet while the occasional arbitrary exercise of editorial power is real, a much more common issue is that finding arm's-length peers is not easy. Given the limits of their own knowledge, editors must sometimes (perhaps often) draw on lists of potential referees submitted by authors themselves, and they lack objective means to learn about authors' personal connections to those referees. Further, the best-qualified referees are often in high demand and unavailable. In such cases, editors may seek referees at some remove from the specific focus area, or rely on more-available junior scholars. In both cases, reviewers' expertise may be insufficient to detect key problems.

Starting in the 1990s, Internet-based publishing opened the door to new models of peer review, including much broader participation. Pre-print servers such as ArXiv (founded 1991) and the Social Science Research Network (SSRN, founded 1994) allowed authors to post draft articles, in part to seek informal commentary,

but also to stake priority claims. In the 2000s, a sea change toward greater transparency across the sciences led to considerable revision of previous norms (Wilkinson et al., 2016). Some journals adopted more open or even fully public review processes, presenting articles online for comment by a larger scientific community, or by anyone at all, before publication. The IPCC has followed this trend.

### 11.3 What Is the Value of Peer Review?

In my own experience as an author, peer reviewer and editor, the process *usually* improves the quality of publication and weeds out many errors of fact, logic and calculation. Yet as suggested earlier, peer review is not a formal audit, its quality is highly variable, it cannot be standardised, it can reflect numerous biases, and it can miscarry, rejecting valuable contributions while accepting shoddy ones. Thus, although scientists hold the practice in high esteem, peer review is anything but a truth machine. So what are its benefits?

First, it operationalises crucial scientific norms. One of these is *methodological scepticism*: peer review invites an evidence-based, ‘prove it to me’ approach to knowledge claims – perhaps the most fundamental element of any scientific method. Reviewing others’ work through this lens teaches reviewers how to think sceptically about their own work as well. Another is *communalism*. Science is organised community learning, a collective effort whose unique value stems from the care and attention of many individuals and the wide sharing of knowledge. Peer review also acts as a form of *expert certification*, similar to advanced academic degrees (reflecting training) and institutional affiliation (reflecting acceptance by other scientists).

Second, peer review serves a *gatekeeping* function. As already observed, this can be highly problematic. Yet it also benefits the scientific community in numerous ways. It reduces the likelihood of error and promotes collective attention to methodology. It also slows growth in the sheer number of scientific publications, a problem in its own right that is now especially acute in climate science (Haunschild et al., 2016). The gatekeeping function of journal review plays a critical role in IPCC assessments, by screening out material self-published by individuals, political interest groups, advocacy organisations, and others. The AR6 WGI report cited over 14,000 publications. Without the gatekeeping role of journal peer review, an almost unimaginable volume of dubious material from websites, self-published books and other ‘alternative’ publication venues might be submitted for formal assessment. This is not speculation: some reviewers of AR6 presented blogs, personal ‘audits’ and other self-published, unreviewed work for consideration in the assessment.

## 11.4 Review of IPCC Assessments

IPCC rules of procedure developed in tandem with the composition of its First Assessment Report (AR1) in 1990. Bert Bolin, the IPCC's first chairman, attached great importance to basing AR1 only on peer-reviewed publications (Bolin, 2007). Peer review of the assessments themselves was discussed at the first session of the IPCC Bureau in 1989, which took the decision to establish a review process that would include scientists from developing countries (Agrawala, 1998b). Importantly, review of science *assessments* differs substantially from journal peer review. Whereas journal reviewers have the power to recommend rejection, IPCC reviewers can only recommend revisions (including elimination of statements or entire topics). The focus of assessment review is therefore to ensure consideration of all relevant material and accurate characterisation of the full range of results (Oppenheimer et al., 2019). Box 11.1 summarises the different forms of peer review conducted by the IPCC, and these are elaborated in the following paragraphs.

**Internal review.** IPCC assessments begin with an onboarding meeting. There, each chapter team begins to fill in and expand the very brief chapter outline previously scoped out by IPCC leadership (see **Chapter 3**). In a few weeks, each chapter team rapidly composes a 'Zero Order Draft' (ZOD). The ZOD is incomplete and quite rough, with many elements existing only as placeholders. The purpose of this stage is to generate a skeleton structure, allow all LAs to get a sense of the entire report, and discover areas where additional content, expertise and cross-chapter interaction will be needed (see **Chapter 18**). LAs comment on the ZOD in a spreadsheet; once compiled, all comments are made available to all LAs. This internal peer review strongly guides early revision.

**Expert review.** Revision of the messy, incomplete ZOD results in the much more developed 'First Order Draft' (FOD), which is then opened to expert review.

### Box 11.1

#### Types and stages of review/scrutiny in IPCC reports

*Journal review.* IPCC reports are based primarily on published, peer-reviewed scientific literature.

*Internal review.* IPCC Lead Authors review their own drafts at every stage.

*Expert review.* Review by scientists and self-declared experts outside the IPCC, starting with the first complete draft.

*Government review.* Representatives of IPCC member governments review middle- and end-stage drafts.

*Approval.* At a final meeting, government representatives approve the Summary for Policymakers (SPM) line by line.

Unlike journal peer review, where journal editors determine who qualifies as a ‘peer’, IPCC ‘expert’ review is open to essentially anyone: ‘Because the aim of the expert review is to get the widest possible participation and broadest possible expertise, those who register are accepted unless they fail to demonstrate any relevant qualification’ (IPCC, 2020a). Despite significant outreach by the IPCC, the majority of reviewers are male and most are from the developed world (see **Chapter 7**).

Most reviewers of the AR6 WGI FOD were climate scientists or others with genuine expertise. However, some very active reviewers listed no affiliation with any scientific organisation and had no publications other than blog posts or other self-published materials. Nonetheless, at least in my own experience, these unaffiliated reviewers occasionally flagged significant errors and contributed valuable revisions. A further observation is that because reviewers’ names are attached to comments, those of senior scientists and experienced IPCC authors may be weighted more heavily by chapter teams. Thus prestige as well as expertise can affect responses to review comments; often there is no principled way to tell the difference. At this stage and beyond, chapter authors are required to respond to *all* comments. If they reject a comment, they must explain why. Typical reasons for rejection include out of scope (for example, promoting a policy, or unrelated to WGI purposes), not supported by published peer-reviewed literature, or no scientific evidence provided.

For authors new to the IPCC – as were about 30 per cent of the 234 LAs, including me, contributing to the AR6 WGI report – the scale of effort required by this review process comes as a very rude shock. It takes approximately four months for the IPCC’s Technical Support Unit (TSU) to format and distribute the FOD for an eight-week comment period, and then compile the comments received. Meanwhile, revision of the draft continues at a rapid pace. This time lag means that chapter text has already been extensively changed and a great deal of new material added before LAs can even start to respond. As a result, responding to comments entails a tedious, confusing back-and-forth between the comment sheet, the formatted FOD and the active working draft.

Despite the warnings of experienced LAs, many of us underestimated the huge amount of time required to do this job well. Many comments cited publications we had not yet considered, requiring us to locate and read them on the fly, or to consult LAs from other chapters for help in interpreting what we learned. Notwithstanding its somewhat chaotic character, this review dramatically improved the draft and extended its evidence base.

**Governmental and expert review.** Revisions to the FOD result in the ‘Second Order Draft’ (SOD). This time, both experts and the 190+ United Nations member governments participate in the review. To avoid politicisation, government

representatives cannot draft any part of the main report; they participate in review on the same basis as experts. At this point several Review Editors – senior scientists with previous IPCC experience – are assigned to each chapter, to provide external oversight of the final review stages. One Review Editor assigned to my own chapter was exceptionally diligent, while the other two were less so. For them as for us, the task of reviewing a 100,000-word, highly technical chapter and evaluating thousands of comments while also working a day job proved overwhelming.

Revision of the SOD leads to the ‘Final Government Distribution’ (FGD) of the Final Draft. At this stage the draft is essentially frozen; however, the TSU revisited comments on the FOD and SOD and required all chapters to respond to any comments they had previously missed or deferred for later action.

**Approval.** In the last 18 months or so of the assessment, each WG designates a subset of authors (Coordinating Lead Authors or LAs) to draft an SPM, typically around 30 pages in length. The SPM is first reviewed by experts and governments, then revised, then subjected to another round of government review. Once finalised, the SPM goes to a plenary approval session, where government representatives approve the SPM line by line.

The role of government representatives is problematic with respect to the concept of ‘peer’ review. While some are very well informed on the scientific issues, others are not, and all are by definition representing the interests of their own nations. The approval session includes both SPM authors (IPCC scientists) and government representatives. IPCC procedures codify that SPM approval ‘signifies that it is consistent with the factual material contained in the full scientific, technical and socio-economic Assessment or Special Report accepted by the Working Group’ (IPCC, 2013a). However, there are many ways to summarise any large, complex document, and seemingly minute changes in language can matter greatly to policymakers’ reception of IPCC reports. As a result, during the approval process government representatives may propose alterations to SPM statements that suit their own purposes (De Pryck, 2021a). Still, the consensus requirement generally limits the power of any one nation in the approval process, and government acceptance greatly strengthens the political authority of the assessments (see **Chapter 20**).

### 11.5 Controversies Surrounding the IPCC Review Process

When AR2 was released in 1996, the IPCC’s rules of procedure became the flashpoint of an intense public controversy. WGI’s SPM and Chapter 8 both included the following sentences: ‘Our ability to quantify the human influence on global climate is currently limited ... *Nevertheless, the balance of evidence*

*suggests that there is a discernible human influence on global climate*' (IPCC, 1996: 5, italics added).

Here, the IPCC for the first time acknowledged a better-than-even likelihood of anthropogenic causes for observed global climate change. The sentence was introduced into Chapter 8 and the SPM by Chapter 8 Coordinating Lead Author Ben Santer following the final IPCC WGI plenary meeting at Madrid in November 1995. There, the exact wording of that sentence was intensely debated – with representatives of some oil-producing states, notably Saudi Arabia, seeking to soften its terms – before the final revision quoted above was approved (Houghton, 2008).

Following release of the revised text, physicist Frederick Seitz and others charged the IPCC with 'deception', saying it had 'corrupted the peer review process' and violated its own rules of procedure (Lahsen, 1999; Oreskes & Conway, 2010). These charges were demonstrably untrue; the changes were introduced by consensus among the participating governments. Nonetheless, the episode drew attention to IPCC rules, which lacked clear closure mechanisms for the review process (Agrawala, 1998b: 624; Edwards & Schneider, 2001). As a result, in 1999 the IPCC revised its rules of procedure and added the Review Editor oversight role described earlier (Skodvin, 2000a; Siebenhüner, 2002).

A second example resulted from controversy over errors found in AR4 (O'Reilly, 2015) and criticism of the IPCC resulting from the 2009 Climategate episode. In 2010, the UN Secretary-General and IPCC Chair jointly requested an independent review of IPCC rules and procedures – including its peer review practices – by the InterAcademy Council (IAC), which appointed a panel of distinguished scientists (see **Chapter 6**). Like many independent commentators (Jasanoff, 2010a; Beck, 2012), the IAC panel found that due to the social significance of climate change and the authority attached to the IPCC's conclusions, 'accountability and transparency must be considered a growing obligation' (IAC, 2010: viii).

The IAC review found the IPCC's existing peer review process essentially 'sound'. However, it noted that the number of review comments had more than doubled, to more than 90,000 for the entire AR4. Fourteen years later, some 78,000 comments were received on the AR6 WGI report *alone*. Adding the comments received by WGII (62,418) and WGIII (59,212), this makes a total of 199,630 comments! The IAC concluded that under time pressure, some review comments might not receive sufficient attention, which is consistent with my own experience.

## 11.6 Achievements and Challenges

The current IPCC review process is the most extensive, open and inclusive in the history of science – a landmark achievement by any measure. Further, the organisation has responded to ongoing critiques with ever greater transparency and

accountability. Today's review process is essentially public, open to anyone (within limits: for example, the English language standard presents a significant hurdle for non-speakers). Since AR4 (2007), the IPCC has published the FOD and SOD of each report on its website, along with all comments and responses. This review process means that minority views and outlier results have been carefully considered by the climate science community at several points, from journal peer review through multiple rounds of assessment review. Nonetheless, no review process can eliminate all errors or guarantee the truth of conclusions.

One very difficult challenge is that comments that dispute the fundamental framing of particular issues may be dismissed, unless a significant constituency supports reframing them (O'Reilly et al., 2012). For example, during review of the IPCC Special Report on Global Warming of 1.5 °C (2018), many commentators expressed 'unease' about the report's presentation of bioenergy with carbon capture and storage (BECCS) as 'a viable carbon dioxide removal technology at grand scale' (Hansson et al., 2021: 1). Yet this misleading framing remained in the final report. Hansson et al. identified several 'boundary work' strategies successfully used by LAs to deflect reviewer critiques of BECCS's potential. For example, LAs claimed that the IPCC mandate restricted them from being 'policy prescriptive' (see **Chapter 21**) – a deflection I encountered and resisted, yet also sometimes used myself, in working on AR6 WGI.

Two further challenges lie in the inexorably growing numbers of relevant publications and review comments. Machine learning techniques have been proposed to augment human processing of scientific literature (Callaghan et al., 2021), but such methods may never be accepted as substitutes for expert judgement. The huge number of review comments already imposes an infelicitous trade-off on volunteer LAs, who must balance their time between careful evaluation of the scientific literature, composition of the report, and responding with care to peer review. Any attempt to restrict the openness of the review process – for example, by requiring reviewers to provide stronger evidence of expertise – could lead to backlash over transparency. Increasing the number of LAs and/or Review Editors might help, yet would also add complexity to an already elaborate report-writing process.

### Three Key Readings

Oppenheimer, M., Oreskes, N., Jamieson, D., et al. (2019). *Discerning Experts: The Practices of Scientific Assessment for Environmental Policy*. Chicago: University of Chicago Press.

This book critically examines practices used by several science assessments, including the IPCC's peer review processes.



Hansson, A., Anshelm, J., Fridal, M., and Haikola, S. (2021). Boundary work and interpretations in the IPCC review process of the role of bioenergy with carbon capture and Storage (BECCS) in limiting global warming to 1.5°C. *Frontiers in Climate*, 3: 643224. <http://doi.org/10.3389/fclim.2021.643224>

This article is one of the few close studies of IPCC peer review of a particular issue.

InterAcademy Council (2010). *Climate Change Assessments: Review of the Processes and Procedures of the IPCC*. Amsterdam: InterAcademy Council. Available at: [https://archive.ipcc.ch/pdf/IAC\\_report/IAC%20Report.pdf](https://archive.ipcc.ch/pdf/IAC_report/IAC%20Report.pdf)

The InterAcademy Council report closely examines all aspects of review in IPCC reports. The changes it recommended have been adopted.