Assessing Bias Charges against Collaborative Expertise, with an Application to the IPCC

William Rehg
Saint Louis University, rehgs@slu.edu

Follow this and additional works at: https://lib.dr.iastate.edu/sciencecommunication

Part of the Other Rhetoric and Composition Commons, Rhetoric Commons, and the Speech and Rhetorical Studies Commons


This Event is brought to you for free and open access by the Conferences and Symposia at Iowa State University Digital Repository. It has been accepted for inclusion in Iowa State University Summer Symposium on Science Communication by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Assessing Bias Charges against Collaborative Expertise, with an Application to the IPCC

WILLIAM REHG

Department of Philosophy
Saint Louis University
3800 Lindell Blvd., St. Louis, MO 63108-3414
USA
rehgsp@slu.edu

ABSTRACT: In controversial science-intensive policy debates, charges of expert bias often arise. How does one sort out such charges—especially when expertise is interdisciplinary and collaborative? In this paper I address the problem of collaborative expert bias at the level of group process. Identification of bias is complicated not only by interdisciplinary complexity, but also by the ubiquity of bias, some of which can be fruitful for scientific discovery. Drawing on the Intergovernmental Panel on Climate Change (IPCC) for illustration, I distinguish different kinds of group-level bias in the sciences and propose ways of identifying bad bias.

KEYWORDS: bias, expertise, collaboration, climate change, IPCC.

1. INTRODUCTION

When scientific experts find themselves entangled in controversial policy debates, charges of bias often arise. How does one sort out such charges—especially when expertise is interdisciplinary and collaborative? In such cases, the alleged bias often operates at the group level. The idea of group bias is hardly new. A substantial literature on group bias now encompasses a range of disciplines and approaches, including cognitive and social psychology, feminism, and science studies. Nonetheless, treatments of bias by informal logicians and critical thinking scholars have focused mainly on bias as a problem for individual arguers. To be sure, scholars of argument have long recognized how group membership can bias an arguer in a particular direction. But the “bearer” of bias remains the individual. Policy-relevant expertise raises the question of bias at the level of the group itself.

This paper addresses the problem of collaborative expert bias in contexts that meet three conditions: (i) a group of experts deliver advice on policy-relevant natural-scientific or health questions that admit of objectively correct and incorrect answers; (ii) reliable answers to such questions require the input of multiple scientific disciplines; (iii) the alleged bias at issue lies at the level of group process. Assessing group-level bias charges becomes especially important in such contexts, given that complex collaborative expert arguments outstrip the competence of any single person to assess. Thus our confidence in the quality of expert advice rests largely on the quality of the process itself.

But how should we make that assessment, given the ubiquity of bias and complexity of technical content? For bias occurs in different forms, and not all bias is bad. After some initial orientation, I distinguish two kinds of group-level bias (secs. 2, 3). I then describe two apparently ubiquitous biases in science: confirmation bias and “preference” bias. Because these can be fruitful for research, it is important to develop criteria for sorting good from bad cases.
To illustrate these issues, I refer throughout to the Intergovernmental Panel on Climate Change (IPCC), and I close with a particular bias charge against that panel (sec. 5).

2. ORIENTATION

The literature on group bias is both sizable and multi-disciplinary, and so some initial sorting is necessary to delimit the scope of my analysis. Notice first the difference between two uses of the word “bias,” which we might label “technical” and “partisan” (cf. Walton, 1999, p. 224ff). In some fields the “bias” label refers to technical probabilistic and statistical errors of reasoning. Statistical biases involve errors in data sampling. Cognitive psychologists use the term “cognitive bias” for certain types of widespread errors in human reasoning, such as the “base-rate fallacy,” and errors in conditional reasoning (Evans & Over, 1996). Statistical and cognitive biases reflect tendencies in human cognition that make certain kinds of technical mistakes likely. Understood merely as a tendency to mistaken reasoning, or as the mistake itself, such biases do not stem from a flawed character or a partisan commitment to defending one’s position at all costs. In this paper I am not so concerned with basic technical mistakes, as with bias in the partisan sense. (To be sure, a partisan bias in disposition can lead to sloppy reasoning, thus to technical biases in content.)

Whereas technical bias appears primarily in the content of an argument, partisan bias can appear in both content and process. At the level of content, Douglas Walton defines partisan bias “simply as a one-sided argument—an argument that lacks the balance necessary for it to be two-sided” (Walton, 1999, p. 76). This definition is helpful for its breadth, capturing a wide range of cases of bias. Indeed, partisanship implies a kind of lack of balance, which might or might not be appropriate. In agonistic settings, such as law courts, we expect each side to produce an unbalanced argument for its position. Balance is achieved through the division of labor. Thus a biased argument is not necessarily a bad argument in context.

The level of process potentially includes everything in the social context that affects the quality of argument-making: the participants’ background and dispositions, the formal and informal procedures they employ, how their personalities interact, the institutional setting, and so on. I am concerned here with group settings, more precisely with “transactional contexts,” that is, the contexts in which members of a group are engaged in person-to-person exchange of arguments. For present purposes, two main aspects of transactional contexts are crucial for the analysis of bias: the participants’ dispositions and the structure of their transaction.

The literature on bias tends to identify partisan bias in process with the arguer’s disposition. Insofar as argument-making presupposes a personal disposition to favor one position over another, hence a kind of partisanship, argumentation seems to require bias. To that extent, we may regard bias as normal or good (Blair, 1988). Walton concurs: “Bias or dialectical slanting in argumentation . . . is not inherently bad”; rather, “if you equate it with advocacy, partisanship, or point of view in argumentation, it may be, in many instances, a good thing” (Walton, 1999, p. xviii). Bad bias occurs when the commitment necessary to engaged argument-making undermines the quality of the transaction or the argument content in context. Although one might associate bias with the arguer’s close-mindedness or lack of impartiality, in some contexts these do not undermine the transactional goal. Thus, a public debate between two close-minded opponents might prove helpful for an audience trying to make up its mind.
3. GROUP-LEVEL PROCESS BIASES

At the group level we can identity distinctively group-level dispositions and procedures as potential sources of bias. To my knowledge, neither of these has received much attention from argumentation theorists.

3.1 Group Dispositions

The prime example of a group-level disposition in the psychological sense is “groupthink.” This phenomenon, identified by Irving Janis (1972), has received considerable attention from social psychologists, though doubts remain about its empirical support, scope, and usefulness as an explanatory model (Esser, 1998; Turner & Pratkanis, 1998). According to Baron’s “ubiquity model,” however, groupthink symptoms—suppression of dissent in groups, tendency to conformity (polarization of attitudes in a group), self-censorship, and the illusion of consensus—typically arise in groups that meet three antecedent conditions: members identify with their group; group discussion gives rise to attitudes that have a normative character for members; and the “situational self-efficacy” of individual members—their confidence in dealing with group tasks effectively—is low (Baron, 2005, pp. 238–244).

Groupthink presents a potentially serious problem in science. Scientific specialties exhibit at least two of Baron’s antecedent conditions. Both empirically informed social studies of science, as well as research on collective intentionality, suggest that (a) scientific cooperation depends on some level of group identification (understanding oneself as a competent member of a discipline), and (b) scientific discussion presupposes and generates empirical and theoretical results that count as normative standards for competent reasoning. When the group faces problems that exhibit the daunting complexity of climate science, we might expect the third condition to appear as well, low situational self-efficacy.

As it turns out, recent IPCC procedural revisions reveal a concern for groupthink in author teams:

Be aware of a tendency for a group to converge on an expressed view and become overconfident in it. Views and estimates can also become anchored on previous versions or values to a greater extent than is justified. One possible way to avoid this would be to ask each member of the author team to write down his or her individual assessments of the level of uncertainty before entering into a group discussion. If this is not done before group discussion, important views may be inadequately discussed and assessed ranges of uncertainty may be overly narrow. Recognize when individual views are adjusting as a result of group interactions and allow adequate time for such changes in viewpoint to be reviewed. (IPCC, 2010, Appendix 4, Treatment of Uncertainty, no. 3, draft)

This statement responds to recommendations from the InterAcademy Council (IAC, 2010) regarding the estimation of uncertainty—more on which below.

3.2 Procedures

In considering group-level sources of expert bias, we must also look to the design and execution of committee procedures. Agencies responsible for expert advice pay close attention

---

1 For relevant work in the social studies of science, see Kuhn (1996); Ziman (1968); for worries about groupthink based on collective intentionality, see Tollefsen (2006).
to committee procedures, whose design and psychology is now the subject of a growing empirical literature (e.g., Bijker, Bal, & Hendriks, 2009; Hilgartner, 2000; Jasanoff, 2005). What we want, presumably, are committees that

- do not dissolve into contention but reach robust conclusions,
- which are scientifically accurate or reliable,
- sufficiently clear to guide policymakers (rather than so hedged with second-guessing as to provide no support for decisions),
- democratically accountable,
- yet not unduly politicized.²

Behind the last condition lies the idea that reliable science for policy should not substantively depend on political partisanship, though it might have partisan political implications. After all, the natural and health sciences—my focus in this paper—reach conclusions about an objective reality, knowledge of which is independent of partisan allegiance, right?

Not quite. In the health sciences, conclusions can hardly avoid cultural assumptions about human well-being. Even when dealing with the non-human world, descriptive categories and risk estimates often involve cultural values; when such values are controversial, politicization is practically unavoidable (Pielke, 2007). In such contexts, we do better to understand “unduly politicized” conclusions as conclusions that do not suffer from bad bias—committee dispositions and procedures foster a balanced assessment of the available science, which takes into account different cultural perspectives and values as appropriate. For that to occur, process designs must succeed in combining rigorous technical analysis with deliberative forums that include the full range of stakeholders, expert and lay.³

In any case, bad procedural biases can arise in a number of ways, all of which involve a lack of appropriate balance in committee deliberation. A committee might lack representation from relevant disciplines for the policy-relevant topic. Or it might lack the wider sources of input for reflecting on social and moral values that condition scientific reasoning and judgment at different points (Douglas, 2009).

IPCC procedures incorporate a number of mechanisms that foster wider input and deliberation. The IPCC is designed to include not only all the relevant disciplines but also scientists from different regions of the world. Moreover, government representatives as well as other stakeholders (business interests, NGOs, etc.) are involved at certain stages of the multi-step report process. That process begins with the selection of authors and determination of report scope; moves through the various stages of drafting, reviewing, and revising of reports; and concludes with plenary sessions in which reports are finally accepted, which is to say: accepted both by scientists and governments (see IPCC, 2008; Bolin, 2007).

---

² This list expands on Douglas (2009), p. 134; for a history of science advising, see Douglas (2009), chaps. 2, 7–8.

³ On the analytic-deliberative approach to science advising, see Douglas (2009), chap. 8; note that inclusion need not require sunshine rules for all expert discussion in committee; see Rehg (2009), chap. 8; Bijker et al., (2009).
4. BIASES INHERENT IN SCIENTIFIC ARGUMENTATION

We should not be surprised if scientific argumentation, like argumentation in general, is inherently biased in certain ways. Here I examine two such biases: confirmation bias and “preference bias.” These appear ubiquitous, but they are not always bad. Consequently, critical assessment of scientific bias requires an eye for the bad cases. Both have their most obvious source in dispositions, but at the group level they can also arise through flawed procedures.

4.1 Confirmation Bias.

Psychological research shows that human beings are inherently prone to favor considerations and evidence that support their position, and to disregard or discount counter-evidence (Mercier & Sperber, 2011; for doubts, see Evans & Over, 1996, pp. 103–109). Up to a point, this kind of bias might be fruitful for scientific discovery. As critics of Popper’s falsificationism have pointed out, great discoveries can require perseverance in the face of apparently contrary evidence. To put it in Kuhn’s terms, good science requires an ability to live with a certain level of “anomalies” in one’s approach, on the assumption that one will eventually be able to resolve inconsistencies between one’s hypothesis and other considerations. Contrary empirical measurements, for example, may simply prove mistaken or insufficiently accurate. Confirmation bias is bad, however, when it leads a researcher to cling to a favored hypothesis beyond all reason, or even worse, when it becomes so channeled and rampant in the community that progress falters.

Mercier and Sperber (2011) suggest that confirmation bias has an evolutionary basis. If human reasoning evolved for its use in social processes of argumentation, then we should not be surprised by widespread confirmation bias in individual reasoning. Though often a weakness at the individual level, confirmation bias can strengthen social argument-making inasmuch as (a) each participant strives to convince others by putting forward the best case for his or her own argument, but (b) by doing this together, participants provide a natural check against each other’s confirmation bias.

As with bias in general, then, context matters in distinguishing good from bad confirmation bias. This is especially true for policy-relevant expertise. Because expert knowledge confers confidence in one’s judgments, expertise can actually accentuate confirmation bias (Mercier, 2011). Even if such confidence can foster discovery in research contexts, it poses a serious problem in policy contexts. Experts, after all, are charged with assessing the overall state of the literature, not pursuing their own research agenda. Thus committee venues are especially important for countering such bias. To do so, the committee must be sufficiently heterogeneous with respect to viewpoints on the issue at hand. Thus composition of the committee and the provision for outside review play important roles in countering confirmation bias. At least at first glance, the size of author teams responsible for IPCC chapters, along with review procedures, should provide some check on expert confirmation bias at the level of individual reasoning.

4.2 Preference (Inductive-risk) Bias

Torsten Wilholt (2009) has argued that “preference bias” is ubiquitous in scientific research. Preference bias “occurs when a research result unduly reflects the researchers’ preference for it
over other possible results” (Wilholt, 2009, p. 92). Thus defined, preference bias coincides with bad partisan bias in general. However, Wilholt is after something more specific: the kind of bias that involves “tampering with the balance of inductive risk” (2009, p. 94). Thus the more accurate term is “inductive-risk bias.”

In testing hypotheses, scientists must make value-laden judgments about the relative costs of two kinds of error: not accepting true hypotheses, and accepting false hypotheses. The widely used 95% confidence level reflects a risk-averse valuation: one would prefer to delay acceptance of a true hypothesis rather than advance a false claim. Decisions that flow from such value-judgments have consequences, not only for the researcher’s career, but in some cases for groups affected by the social consequences of the decision—for example, the delay of new treatments for the critically ill. But, Wilholt maintains, there is no principled yet feasible method for impartially balancing the costs and risks involved in the two kinds of error. The reason is that we cannot distinguish honest disagreement in judgments of value from genuinely bad bias. Inductive-risk bias, then, is both ubiquitous and ambiguous—it “will simply be part of the scientific condition” (Wilholt, 2009, p. 95; also Douglas, 2009).

Inductive-risk bias can arise at various stages of research: in experimental design, interpretation of data, and communication of results. Biased experimental design, for example, makes it less likely that one will falsify one’s favored hypothesis. Bias can affect interpretation of data in various ways, such as through one’s choice of statistical standards for rejecting data. Some feminist critiques of masculinist bias in science lie at this level (e.g., Longino, 1990). Finally, publication bias is an example of inductive-risk bias that affects communication of results. Publication bias arises from the fact that researchers and journals systematically prefer positive findings over research that fails to confirm a hypothesis. For example, if climate-journal editors believe the social-environmental risks of underestimating the effects of climate change are greater than the risks of overestimation, they might preferentially publish studies that confirm the threats of climate change. Climate change skeptics have charged climate science with such bias. But without a record of high-quality unpublished findings, the charge is speculative at best (for a reply to skeptics, see Nordhaus, 2012).

Drawing on social epistemology, Wilholt identifies disciplinary conventions as one solution to the problems posed by the ambiguity of inductive-risk bias. The key problem lies in the corrosion of trust necessary for fruitful scientific collaboration. Scientific practices heavily depend on the mutual trust scientists place in their peer-reviewed arguments. Conventional standards in effect remove some judgments about inductive risk from the individual’s purview. The result is a better coordination in the normative standards scientists use in evaluating results. Insofar as their papers adhere to conventional standards of experimental design, confidence levels, and interpretation, differences in individuals’ judgments about the value of this or that hypothesis and relative costs of error have less effect on results, and epistemic trust is preserved across the community. And insofar as unpublished negative research findings are made available for scrutiny (as in some medical fields), suspicions of publication bias can be confirmed or rejected (Wilholt, 2009, p. 97). This analysis leads Wilholt to a usable concept of inductive-risk bias: “the infringement of an explicit or implicit conventional standard of the respective research community in order to increase the likelihood of arriving at a preferred result” (2009, p. 99).

Nonetheless, context matters here too. Inductive-risk bias is not necessarily bad, inasmuch as different risk preferences in the community can feed competing research programs and therewith the exploration of more avenues of discovery. But expert committees pose
distinctive challenges. Similar to scientists’ use of conventional research standards, expert committees, the IPCC included, develop their reports using procedures that are supposed to foster the audience’s trust in the report content (Hilgartner, 2000). Unlike research standards, however, committee procedures are not designed simply to coordinate practices across a discipline by removing value-judgments from individual discretion. Here the difference between long-term and short-term timeframes is decisive. As long as members of a discipline follow the same conventions, whether those conventions favor risk-taking or risk-averse strategies is less important than the fact that they do not squelch discovery and correction of error over time. In science advising, a genuinely impartial balance between risk and caution becomes crucial for the trustworthiness of policy-relevant advice, which is inherently value-laden and thus potentially politically contentious. Consequently, expert procedures must reflect such impartiality, both in the estimation of technical uncertainties and in their responsiveness to democratically accountable deliberation over the relevant values.

I thus suggest that bad inductive-risk bias in science advising occurs when an expert committee infringes publicly accepted, democratically accountable guidelines for assessing relevant literature, estimating uncertainties, and characterizing risk in an impartial manner. Note that this definition does not depend on any attribution of suspect motives to committee members; the bias might involve such dispositions, but it can also arise from unintentional procedural violations. Adequate procedures must ensure both technical reliability and, at some level, democratic accountability—demands that federal agencies and lawmakers have recognized for some time now (Douglas, 2009, chap. 7).

5. BIAS IN THE IPCC

I close by applying the above analysis to an actual bias charge against the IPCC. Critics have leveled a range of bias charges against the IPCC: ignoring relevant literature; groupthink; political bias; and publication bias in climate science as a whole (resp., Pielke & Staley, 2007; Lemonick, 2010; Horner, 2007; Michaels & Balling, 2009, chap. 7). Given the inclusive, multi-stage design of IPCC report-writing procedures, bias charges face a significant burden of proof. In general, a strong bias charge combines multiple indicators of possible bias (Rehg, 2011, pp. 396–397). In the case of the IPCC, the stronger charges link (1) technical problems in report content, (2) flaws in procedures or their execution, and (3) evidence of close-mindedness.

Judith Curry, an active climate researcher, has issued a string of criticisms of the IPCC that link these elements (see Lemonick, 2010; Curry, 2010; Curry, n.d.). What makes her critique interesting for my purposes is her suggested tie between the IPCC’s problematic uncertainty estimates and groupthink bias, more precisely a defensive mentality in the IPCC that cuts off potentially fruitful dialogue with more responsible critics. This charge fits with the contents of the purloined emails of some IPCC scientists, whose remarks indeed display a certain defensiveness (Powell, 2011, chap. 14). Moreover, the IPCC apparently meets the antecedent conditions that make groupthink likely, and both IPCC scientists and critics have concerns about procedural flaws (see IAC, 2010; Hulme, Zorita, Stocker, Price, & Christy, 2010). So we should not simply dismiss Curry’s critique.

But we should hesitate before accepting Curry’s defensiveness charge. The IPCC procedural revisions reflect a concern not with defensiveness, but with factual errors and group dynamics that engender a kind of team-level confirmation bias. Nor is it clear that IPCC
scientists are defensive about their process. Besides accepting certain IAC recommendations (IPCC, 2010), IPCC scientists have engaged criticisms online (at realclimate.org), and have publicly entertained a range of structural revisions to the IPCC (Hulme et al., 2010).

What is more, both Curry’s technical objections about uncertainty and her defensiveness charge fall under the ambiguous category of inductive-risk bias. We should thus ask, first, to what extent the technical objections simply express different value-judgments of the costs of different kinds of error. This question lies at the heart of the political controversy over climate projections. Before we can connect uncertainty estimates with bad bias, we must clarify the values that underlie competing assessments of the possible impacts of climate change and response strategies on the economy, environment, and public health.

Second, we should ask to what extent the defensiveness charge boils down to an honest disagreement over the merits of two transactional strategies. Some of Curry’s own remarks support this reading:

> There seems to be some sort of unwritten rule by the IPCC scientists and their defenders not to engage with critics/skeptics, since they think that such engagement legitimates the skeptics. Personally, I think that the almost total lack of “mainstream” climate scientists engaging with skeptics has resulted in a loss of the moral high ground in the public’s view, and has acted to increase the public credibility of the skeptics. (2010)

At this point we do not have a straightforward bias charge, but a genuine dispute over transactional merits. At issue are the proper conduct of science advisory panels and their mode of public engagement.

More precisely, we have two procedural alternatives—engaging critics/skeptics more widely or more narrowly—that incorporate different value-judgments about the costs of different kinds of error. As some analysts have pointed out, opening modeling data to critical scrutiny can assist in the correction of errors (Edwards, 2010, pp. 421–427). However, too wide an engagement with skeptics can retard argument-making—in effect, one raises the bar on acceptable conclusions, similar to raising the confidence level for acceptable hypotheses. If one judges the risks of inaction as the more serious (compared to the risk of unnecessarily aggressive measures), then a more controlled admission of criticism has some justification.

The narrower strategy has its roots in the history of climate-change skepticism, whose ties to business interests are well-documented (e.g., Oreskes & Conway, 2010; Lahsen, 2008). I think the record shows that the IPCC did not initially dismiss industry-sponsored and libertarian skeptics; rather, scientists were unconvinced by skeptical doubts, and as typically happens in science, unconvinced scientists eventually stop listening to holdouts. The field moves on, presuming that the truth will prove itself at the level of practice, in the comparative fruitfulness of competing research programs. The failure of skeptics to display the appropriate ethos of scientific argumentation only hastened their dismissal: many of them lacked credentials in climate science, and others argued in a way that points to bad faith (see Oreskes & Conway, 2010). Moreover, skeptics often tend to repeat old objections to climate science, but without any mention of the strong rebuttals on record. These ethotic failures partly stem from differences between the ethoi of political and scientific argumentation. Because politics, more than science, is characterized by mistrust, stricter adherence to predetermined procedures becomes crucial to legitimate outcomes, something IPCC scientists did not seem fully to

---

4 I have noticed this in my own research on the history of the controversy; I also draw here on conversations with a climate scientist, Benjamin de Foy. But see also Powell (2011).
appreciate (see Edwards & Schneider, 2001); In addition, the self-expressive character of political argument allows losing parties to repeat their position far longer than would be appropriate in science.

In support of the more open approach, Curry points out that climate-change skepticism has evolved over the last decade from industry-sponsored attacks to decentered, blog-based challenges that do not stem from special interests (Curry, n.d.). Thus climate scientists’ wariness toward outsider critique no longer obviously holds. There may be advantages, both for the public credibility of climate science and for its technical merits, in developing some wider venues of public engagement that could positively interact with expert forums such as the IPCC.

However one settles this question of procedure, one should do so in the honest recognition that in the United States, the skeptics have won the rhetorical contest over inductive-risk bias, for they have cast sufficient doubt on the impartiality of IPCC process to undermine trust where it matters most: in the halls of Congress and in the portion of the population that supports skeptical legislators. In winning this victory, skeptics were assisted by well-publicized infringements of IPCC procedures designed to safeguard impartiality. It is time for public reflection on how to improve the communicative dimensions of IPCC process.

REFERENCES


