Canadian Journal of Philosophy (2024), 1–14
doi:10.1017/can.2024.25

ARTICLE

No Peeking: Peer Review and Presumptive Blinding

Nathan Ballantyne and Jared Celniker

School of Historical, Philosophical, and Religious Studies, Arizona State University, Tempe, AZ, USA
Corresponding author: Nathan Ballantyne; Email: nballantyne@asu.edu

Abstract
Blind review is ubiquitous in contemporary science, but there is no consensus among stakeholders and researchers about when or how much or why blind review should be done. In this essay, we explain why blinding enhances the impartiality and credibility of science while also defending a norm according to which blind review is a baseline presumption in scientific peer review.

Keywords: peer review; research evaluation; biases; open science; brutal humility

1. Introduction

Science is a social enterprise in which human beings hypothesize, collect data, and draw conclusions. But what happens in the laboratory does not always stay in the laboratory—science also involves the transmission of ideas and knowledge out into the scientific world and perhaps then into further communities of journalists, policymakers, and the general public. Since the early modern period, scientific ideas have standardly been distributed through scholarly publication (Biagioli, 2002; Csiszar, 2018) and peer review has become one important aspect of the process that determines what reaches the scientific public (Lee et al., 2013; Weller, 2001).

We are interested in a relatively neglected aspect of the review process: the "blinded" evaluation of research. As we use the term, blinding involves removing or masking certain information when a reviewer evaluates research—removing what we call *task-irrelevant information*. When blinding is successful, reviewers only have access to *task-relevant information* (Thompson, 2016). Most commonly, blinding involves removing or masking details that could enable reviewers to identify authors, but, as we will explain, sometimes task-irrelevant information also includes data and interpretations of data, when a reviewer is tasked with evaluating a study's methods.

The terms "single-blind," "double-blind," and "triple-blind" are ubiquitous in contemporary science, but there is no consensus among stakeholders and researchers about when or how much or why blind review should be done. In this essay, we will defend the view that, simply put, blinding is good.

We understand the goodness of blinding as follows: (1) blinding is *necessary* for good science (thought of as a broad social enterprise) in virtue of facts about human psychology and social relations; (2) blinding is *never sufficient* for good science because good science also depends on background conditions such as researchers' and evaluators' knowledge, skills, and resources; (3) in practical contexts, challenges to blinding—in the form of conflicts between blinding practices and the values held by scientists—tend to be merely apparent and often can be addressed with minor tweaks to the review process; and (4) more blinding is usually preferable to less.

Why does it matter that blinding is good? Briefly, people ought to promote good things when they can do so without thereby incurring undue costs. And we believe that blinding can be easily and

© The Author(s), 2024. Published by Cambridge University Press on behalf of The Canadian Journal of Philosophy, Inc. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

without undue cost expanded in scientific review. What's missing from discussions of blinding, though, is an account of when, how much, and why blinding should be done. For instance, scanning across journals from various disciplines, we find a surprising variety of processes and norms governing peer review. Consider a pair of examples. One of the most prominent science publications in the United States, *Proceedings of the National Academy of Sciences* (*PNAS*), features single-blind review, where authors do not know their editors or reviewers but both editors and reviewers know the authors' identities. In the recent past, *PNAS* even offered a "contributor track" where members—that is, elite researchers elected to the National Academy of Sciences—could choose their own reviewers for their submissions (Aldhous, 2014). Nice work if you can get it. Some other publications' blinding policies are more stringent. An academic philosophy journal, *Ergo*, practices triple-blind review, where authors' identities are unknown to both reviewers and editors, and authors do not know reviewers' identities. In emails to potential reviewers for manuscripts, *Ergo*'s editors provide instructions meant to avoid inadvertent unblinding:

> If you know the identity of the author, please decline *without explanation*. In particular, please do not say that you know the identity of the author, as this may reveal it to the editor sending this request. Ergo uses triple-anonymous review, so it's even possible that you are the author, unbeknownst to the editor. Again, if this is the case, *please do not say so*.

Suffice it to say, blinding is taken with varying degrees of seriousness at different publications and in different fields.

Our goal in this essay is to articulate and defend a plausible account of blinding—one that helps explain when, how much, and why blinding should be done—so that stakeholders and researchers can know how to more effectively implement blinding in their review processes. Science needs norms to guide behavior and coordinate collective efforts, and implementing better norms about blinding could improve the quality of scientific inquiry.

We proceed by answering three questions:

What is a blind review?

What are its benefits?

What are its downsides?

Then we conclude with some questions about the future of blinding in an increasingly open and interconnected scientific world.

2. What Is Blind Review?

Blinding is always relative to an evaluation or judgment task. Information that needs blinding in one context might be information that is essential for making a proper evaluation in another. (We will return to that point about the relevance of context.) What calls for some information to be blinded is the requirement or obligation or goal to do the task in a certain way—namely, in a way that is impartial, unbiased, or fair. Take a commonplace example to illustrate. In a long-running television series, *The Voice*, aspiring singers initially perform for celebrity judges in "blind auditions." When each singer performs, the judges turn away from the stage. The relevant information is exclusively the quality of the singing. At least momentarily, the singers' physical appearance and mannerisms are treated as irrelevant in the competition.

When reviewers and editors evaluate scientific research, there are two types of information that could be treated as task-irrelevant. First, there are facts about an author's identity: name,

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

professional affiliation, gender, race, age, socioeconomic status, and professional track record. We call that *authorial information.* Second, there are certain facts about the data. This includes information about the data itself, its collection, and its interpretation.

We should emphasize why each type of information might be helpful to eliminate during peer review. First, the psychosocial dynamics between authors and evaluators can be powerful, potentially skewing the process of evaluation. Friendship, cronyism, and nepotism can make reviewers less critical than they would be otherwise. But scientific review also happens amidst all of the seething, subterranean energy of human personality—competition, prestige, jealousy, grievance, and animosity. In 1845, an article in a London magazine described scientific referees as quite possibly "full of envy, hatred, malice, and all uncharitableness" toward authors (quoted in Csiszar, 2016, p. 308). Whether it is over scientific ideas or anything else, people judging people can be ugly. Second, factors concerning the work itself can trigger evaluators' prejudices, including the political implications of the findings, the relevance of the findings for evaluators' theoretical commitments, and so on.[1]

As we noted, blinding is relative to an evaluation. Importantly, a reviewer could be assigned many types of evaluative tasks. We find it helpful to gesture at two categories here: *vigilance* and *interestingness tasks.* On one hand, vigilance tasks require determining whether a manuscript's methods and arguments are appropriately aimed at truth. Consider several questions a journal editor could ask evaluators about a manuscript: Is the literature review adequate and accurate? Are the methods suitable for addressing the research question? Are the interpretations of the data based on good evidence and reasoning? What we call interestingness tasks, on the other hand, require determining how well a manuscript is suited to contribute to an ecosystem of research, beyond the mere truth value of the manuscript's claims. Questions for reviewers might include the following: How novel are the results? Does the data advance theory or practice in a field? How much uptake will there be among researchers or other audiences? It is important to recognize that information that is necessary for completing one type of task might be unnecessary for completing another.

Our interest here is blinding in scientific review, but blinding is used elsewhere as a tool for improving human judgment. For instance, in legal proceedings, fingerprint analysts are sometimes blinded from information about the defendant's alibi, in order to reduce potential biases in fingerprint analysis. What information is relevant or irrelevant to the task shapes what gets blinded or not (Dror et al., 2015; Thompson, 2016).

We propose a stipulative definition of task-irrelevant and task-relevant information. First, information is task-irrelevant when it is not good evidence for accurately answering the question posed by a task, whereas information is task-relevant when it is good evidence for accurately answering the question posed by a task. We assume that evidence is good (or not) insofar as it does (or does not) make someone more justified in accepting an answer to a question. We also note that, according to our definition, a piece of information is either task-irrelevant or task-relevant for a given task, but never both. That follows from the fact that information cannot be both good evidence and not good evidence for answering a question.

In light of this way of thinking about task-relevant and task-irrelevant information, an issue arises: Is task-irrelevant information simply biasing information? After all, discussions of blinding practices often emphasize that an author's identity or name could be biasing (Huber et al., 2022; Largent & Snodgrass, 2016; Weller, 2001, chapter 7). Just for clarity, let's assume that bias is "a systematic departure from a genuine norm or standard of correctness" (Kelly, 2023, p. 4). While it is

[1] Notice something else that follows from our distinction between authorial and non-authorial information. Some academic observers favor "anonymization" over "blinding" language. One reason we use "blinding" talk is that whenever non-authorial information is at issue, it seems unhelpful and confusing to talk about "anonymizing" that type of information—there are no identity features to suppress in such a case. Another reason is that the vast majority of past research about whether peer review should mask certain information uses the term "blind review."

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

clearly true that reviewers could be prejudiced against certain authors, our proposal is that authorial information is typically task-irrelevant because having that information does not provide good evidence for answering a question about, say, the quality of a study's methods or data. Moreover, some information could be task-relevant but at the same time biasing for particular evaluators. Imagine a curmudgeonly, old-school reviewer who is annoyed by new statistical methods and thus negatively biased toward any research using those methods. Then the mere fact that some research uses those methods is, for the old-school reviewer, both task-relevant information and biasing. Unfortunately, providing reviewers with what is task-relevant and shielding them from what is task-irrelevant cannot eliminate all of their biases.

Here is a further question about our definition: Is there a strict, invariable, and transdisciplinary classification of what is task-irrelevant and task-relevant? In other words, supposing some information is task-irrelevant, does that mean that it is always and everywhere task-irrelevant? No. To suggest why, we note two types of examples.

First, certain information might be task-irrelevant in one phase of a review process but become task-relevant later on. To illustrate, consider that reviewers may sometimes evaluate the quality of methods by assessing whether the findings accord with their prior beliefs about the topic at hand. Such evaluations could involve motivated reasoning: if the reviewer has a negative view about a study's results or theoretical implications, their negativity might bleed into how they evaluate the methods (Celniker & Ditto, in press; Cusimano & Lombrozo, 2023). One way to mitigate such biases is to implement "tiered review" and "sequential unmasking" (Button et al., 2016; Dror et al., 2015). In a tiered or sequential review, the reviewer is first asked to evaluate the methods; after they complete that initial task, they are asked to review the data and its interpretation. Temporary blinding and unblinding ensure that the reviewer's impressions of the data and interpretation do not skew the reviewer's evaluation of the methods. It is also worth pointing out structural similarities between tiered review and registered reports. In a registered report, authors submit to a journal a proposal for a study, and the proposal receives an in-principle acceptance (or rejection) solely on the basis of the research question, supporting theory, and empirical methods. Data collection happens only after evaluators decide a proposal adequately addresses the research question (Chambers & Tzavella, 2022). After the study has been run, the data are reviewed to ensure fidelity to the initial proposal and to assess the justification offered for conclusions. What's crucial to notice here is that in the processes of both a tiered review and a registered report, the very same information (e.g., the data) is both task-irrelevant and task-relevant at different stages because the tasks at hand change.

Second, in some disciplines, stakeholders and researchers take for granted that facts about an author's identity should be blinded, but in other fields that judgment will be contested. In 2017, matters of research, identity, and journal review policies erupted in academic philosophy when a white female researcher published in the feminist journal *Hypatia* an article on transracialism, featuring the example of Rachel Dolezal, an American woman born to white parents who claimed to be black (Tuvel, 2017). Although the *Hypatia* affair is complicated and contentious, we note that some critics of the article insisted that the author's identity was relevant to the journal's editorial process. The idea that evaluations of research should take into account a researcher's identity also emerges in fields outside of feminist philosophy. In a fictional example, an ethnomusicologist works on rural Tibetan music. The ethnomusicologist visits Tibetan villages to document how musicians learn, engage with traditions, and perform. Now suppose the researcher is of non-Asian ethnicity. In the ethnomusicology research community, some could worry that a non-Asian researcher's data collection could be skewed by the researcher's identity. Arguably, the race or ethnicity of the researcher is task-relevant in reviewing the study's methods and data. We mention in passing that the recent uptick in "positionality statements," in which researchers disclose their group identities, reflects the fact that identities can sometimes influence how researchers' conceptualize and conduct their work.

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

Our view is that which information is task-relevant and task-irrelevant, and when in the review process they might acquire whichever status, is up for grabs inside a field. Our rationale stems from a truism about evidence: what counts as good evidence for people to believe something always depends on what other evidence is available. To make general proclamations intended to hold for *all fields* seems to overreach because the evidence available across fields could vary significantly. (In fact, we think it might sometimes be difficult to make useful proclamations for a single, broad field of research, in which case the matter is better left up to journals or professional organizations that represent subfields.[2]) But when some facts are determined by researchers and stakeholders to be task-irrelevant, we believe that is a strong reason to conclude they should be blinded.

That claim immediately raises a question: Why? What exactly are the benefits of blind review?

3. What Are the Benefits of Blind Review?

We have a short answer: blind review increases both the impartiality of the review process and the social credibility of the ideas produced by a field. According to us, both impartiality and social credibility benefit science.

Let's say more, starting briefly with social credibility. We agree with Carole Lee and colleagues when they write that "peer review signals to the body politic that the world of science and scholarship takes seriously its social responsibilities as a self-regulating, normatively driven community" (2013, p. 3). We believe that a field's commitment to more blinded peer review even more strongly signals responsible self-regulation than does a commitment to less blinded (or entirely unblinded) review. Just consider a field made up of single-blind journals like *PNAS* to a field with triple-blind journals like *Ergo*. If that is all you know about these two fields, which one strikes you as more credible? We hypothesize that outsiders and non-specialists will place greater trust in a field that they know practices blind review than one that does not, all other things being equal, because demonstrating a commitment to mitigating biases engenders trust. In line with this hypothesis, there is evidence indicating that when people perceive that institutions are biased, they trust them less, even when the bias is toward one's own ideology (Clark et al., 2023).[3]

We also claim that blind review increases the impartiality of review. Notice what that does not mean. Impartial review does not guarantee an accurate review. Even the most impartial reviewer can err. An impartial reviewer could find merit in methodologically flawed research or fail to see the quality of good research. Impartiality does not entail accuracy. It's about treating like cases alike, without preferential treatment or discrimination toward authors as well as toward data and interpretations in a tiered review or registered report. At bottom, what matters in scientific review is task-relevant information. We will say that a reviewer is *impartial* when they make their evaluations by attending to task-relevant, not task-irrelevant, information. Impartiality comes in degrees, to be sure: a reviewer who sticks entirely to the task-relevant information is more impartial than another who takes some account of both types of information. Thinking about impartiality in modal terms is useful: a reviewer is robustly impartial when they would judge each instance of research presented to them by only attending to the task-relevant information.

According to us, blinding is required to encourage impartiality because reviewers can't be expected all on their own to be impartial once they encounter task-irrelevant information. That information cannot improve the quality of their judgments and might even bias their evaluations. Rather than trust that reviewers can ignore or bracket task-irrelevant information, blinding ideally prevents it from playing a role in evaluation in the first place. We see blinding as one practice of

[2] We thank William Sandoval for discussion here.

[3] Is social credibility exclusively beneficial for science due to its influence upon outsiders to a field? We don't think so: a field's members can also benefit from awareness of the types of blind review processes used to evaluate scientific claims. Establishing a field's credibility both within and outside of it is important, though we can imagine different cases for which one benefits science more. We ignore those issues for now.

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

"brutal humility"—where occasionally arrogant, dogmatic, and overconfident thinkers are *forced* to act more humbly. It is not just other people who need brutal humility, of course—sometimes we do not recognize our own biases (Ballantyne, 2015, 2022; Cheek & Pronin, 2022).

Aren't we being too paternalistic here, though? Can't scientists be trusted to have good motivations to base their judgments solely on the task-relevant information, even if they happen to peek at task-irrelevant facts? We think many scientists often do have good motives, but good motives are insufficient to prevent task-irrelevant information from biasing judgment. On that matter, evidence from a recent field experiment is illuminating. Researchers who were invited to review a manuscript were more likely to accept the assignment when a Nobel Prize winner was listed as the corresponding author, compared to when the authorship was anonymized or when the author had low professional status. What's more, reviewers were more likely to recommend accepting the manuscript and less likely to recommend rejection when the author was prominent, compared to the blind or low-status conditions (Huber et al., 2022). The status of the author influenced the reviewers' decisions, and similar effects have been documented in other fields (Fox et al., 2023; Largent & Snodgrass, 2016; Tomkins et al., 2017). Social scientists have shown repeatedly that status biases are commonplace among ordinary people (Cuddy et al., 2008) and, as experimental psychologist Eric Uhlmann noted, "[i]t would be remarkable indeed if scientists were immune to the empirical phenomena [they] study" (Uhlmann, 2011, p. 214).

We said that impartiality does not entail accuracy, but is there some other relationship between impartiality and accuracy? We think so, but it is somewhat indirect. Impartiality makes accuracy more likely over a series of trials because it eliminates certain kinds of error that task-irrelevant information could easily introduce.[4]

But let's pause to note a puzzle here. We say that peer review can become more accurate over time by eliminating certain pieces of information. On first blush, removing information from a decision-making process seems likely to lower accuracy. Why do we think accuracy can be improved by excluding information? Is less really more here?[5]

To dig into the puzzle, consider a case. An author's past performance might sometimes be a strong signal about the quality of their new manuscript. Imagine a reviewer is evaluating a manuscript. While doing the review, they know the author's track record is a better source of evidence about the manuscript than is the reviewer's own assessment based on blind review. For instance, if you know that 95% of an author's past papers are good, and you know that your reliability for discriminating between good and bad papers is lower than that, then trusting the track-record information instead of your blind assessment alone will lead you to better decisions. In this case, isn't blinding a hindrance?

[4] If our suggestion isn't obvious, here is an analogy to explain. The U.S. game show *The Price is Right* has a game called Plinko. The Plinko board features a series of pegs in offset rows; the board is angled upward at close to 90°. At the bottom of the board, there are slots with a range of prize values, from thousands of dollars to zero. On the show, a contestant releases a chip at the top of the board and the falling chip is then deflected by the pegs, making it difficult to predict where the chip will come to rest. The contestant hopes the falling chip will land in a high-value slot, but Plinko is a game of chance and an outcome of zero is not unlikely. We propose that evaluating scientific research is a bit like Plinko. Judging the quality of research is an unpredictable process, involving fallible reviewers and editors. Drop a chip into the board and watch it ricochet to and fro. But in science we try to decrease the kinds of errors task-irrelevant information could induce, thereby enhancing the accuracy of peer review. Blinding does that by blocking some of the least valuable outcomes on the Plinko board. By focusing reviewers' attention on facts that are relevant to the evaluative task and masking those which could not improve judgment and could in fact introduce bias, we can make peer review more accurate by reducing predictable biases.

By invoking a game of chance as a metaphor for peer review, we do not mean to imply that reviewers' skill plays no role. We intend to highlight a salient difference between Plinko and what we could call "rigged" Plinko, where pathways to the lowest-value slots are blocked. Eliminating task-irrelevant information in review is akin to rigged Plinko. The way the game is rigged is analogous to skillful reviewing under conditions of blinding. In general, skillful reviewers might be more accurate than less-skillful reviewers, but skillfulness alone does not entail an ability to avoid all of the biases that can derail productive peer review.

[5] For helpful discussion here, we thank David Christensen, Marcello Di Bello, and Peter Kung.

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

The case draws attention to a different type of error than the one we have focused on so far. We have noted that accuracy can be impeded when information is *available* to evaluators—for instance, authorial information can bias judgment. But accuracy can also be impeded when information is *unavailable*—in situations where authorial track-record information would in fact "enlighten" judgment. We grant that both errors are possible, but we do not view them as equally threatening to scientific review.

While reviewers' evaluations of a study can certainly be biased by authorial information, there is no clear evidence showing that blinding such information impedes reviewers' accuracy. For example, one study examined talks and posters submitted to a meeting of the Society of Judgment and Decision Making (Pleskac et al., 2023). The study revealed that reviewers exhibited biases in single-blind review (e.g., by favoring senior coauthors and disfavoring Asian authors) that did not emerge in double-blind review. The authorial information was prejudicial in this case, but did the prejudices improve the single-blind reviewers' assessments of submission quality? As it turns out, single-blind reviews were no more predictive of outcomes such as expert-rated submission quality, talk attendance, and future publication status than double-blind reviews. In other words, single-blind reviewers demonstrated clear evaluative biases, yet these biases did not enhance reviewers' judgments of quality, compared to double-blinded reviewers'.

Regrettably, there is little data that addresses the matter of whether authorial information in an unblinded review could be enlightening. For theoretical reasons, we doubt such evidence is forthcoming. Again, we grant that, hypothetically, evaluators' reliability could be reduced by blinding. Notice an important limit here, though: we don't think anyone will claim that unblinded review always and everywhere improves accuracy, but only that it does in some cases. Which ones? And how would you know when you are in those cases? As a matter of editorial practice, we don't see a straightforward way to determine when authorial information would be reliably enlightening.

There are several practical challenges for using track-record information. First, it is important to distinguish between *having good track-record information* and *knowing you have good track-record information*. If you merely possess the information but don't recognize its quality, trusting it could be unwise. Second, to discriminate between reliable and unreliable track-record information, you must rely on some rules or heuristics, but it is unclear how to calibrate your judgment. Obviously, a researcher's past work is never *identical* to their new work, in the way one widget produced by a machine is identical to a widget produced last week by the same machine. Many factors might vary in the research process across time, including changes in topic, changes in techniques (data collection and interpretation), changes in the composition of the research team, and changes to the roles within a team. Determining the complex interplay of these factors and appropriately weighting their value is non-trivial. Finally, even supposing you have a way to overcome the previous obstacles to identifying reliable track-record information, eventually, that information's predictive power may decrease. Scientists build their reputations on the basis of past research, but the signal can decay rapidly. They might work hard until tenure or promotion and then lower their standards. As La Rochefoucauld noted, "Men's worth, like fruit, has its season" (Rochefoucauld, 1678/1959, p. 75, §291). Even when you know a researcher's past record is exceptional, how would you know when they have started to grow mold?

But there is another reason we think knowing a researcher's track record is practically infeasible. Although the public record of science is published work, the road to publication is paved with rejection. Not all projects pan out. When a track record is constituted entirely by published work, there is a glaring survivorship bias: some papers did not ever see publication, perhaps because they were poorly conceived or executed. Available track-record information almost certainly will not account for an author's misfires and failures. If you try to infer that an author's work is good on the basis of past good work, but you ignore or can't account for bad work, your inference will be shaky.

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

Let's quickly turn to a worry some people might have.[6] We granted that track-record information could, in theory, have significance for the evaluation of research. Doesn't that commit us to treating track-record information as potentially *task-relevant*? After all, the potential benefits of using track-record information might outweigh the costs of the biases such information could induce. If track-record information is sometimes task-relevant, then by our own lights we would have little reason to advocate blinding it. The worry is ungrounded, though. As we have argued, real-world evaluators will never be in a position to know when their track-record information is a proper basis for judging a new manuscript's quality. Track-record information is not a reliable guide for those evaluations, so we don't treat it as good evidence—it is, on our account, best regarded as task-irrelevant. The theoretical utility of track-record information does not make it a reliable source of information for real-world purposes.

All things considered, the benefits of blinding (and the risks of unblinding) suggest to us there is a *presumption* for blind review. Stakeholders could increase a field's impartiality and social credibility by blinding in review any task-irrelevant information. What we call presumptive blinding is that the benefits of blinding can be assumed to outweigh the costs unless proven otherwise. The idea can be expressed as a norm: when some information is determined to be task-irrelevant, it should be blinded, unless blinded review is determined to have a significant downside. The norm is regulative in spirit. It tells us how science can slouch its way toward accuracy via impartiality without telling us whether any particular scientific claim is accurate or not.

So far we have described some benefits of blinding, but what might its downsides be and are those ever significant enough to override the presumption of blinding? In the next section, we turn to those matters. But first, we want to consider three objections to the norm of presumptive blinding.

The first objection is that publishing practices are diverse and some types of scholarly communication are not typically reviewed with the authors' identity blinded.[7] That being so, the norm must be defective, because it recommends a presumption of blinding in cases where there's a presumption of unblinding.

In response, we agree that particular types of research-related writing do not call for blind review, but that is consistent with the norm: we suggested that names and identities can sometimes be task-relevant facts. We will discuss two relevant examples to clarify our thinking about scholarly publishing that does not report new data or experimental results. First, seasoned researchers are occasionally invited by a journal to reflect on their careers; and researchers sometimes write in an "op-ed" style about a field's current trends or future opportunities. Such papers are, as a matter of fact, frequently reviewed with authorial information unblinded. But we think that if these papers happened to include an experimental study or new data, the presumption of blinding would need to be taken seriously—suggesting the presumption was there in the background all along. So, when such manuscripts are not blinded, we allow that authors' names could well be task-relevant. For a second example, imagine that a Nobel Prize laureate writes a manuscript repudiating her earlier work. Arguably, the Nobel laureate's identity is task-relevant for evaluators. But suppose further that the laureate's arguments against her past research are bad. To be sure, bad arguments published by a prominent researcher could easily waste other scholars' resources as they attempt to refute them. Although a *mere statement of repudiation* could be reviewed without blinding, an *argument for repudiation* might deserve blinded review—but if not, then the editors or stakeholders would need some rationale for treating the author's identity as task-relevant. Our point is that the norm of presumptive blinding provides a baseline for evaluation protocol, even when blinding doesn't happen.

Move to a second objection. In small, tightly knit research fields, reviewers of identity-blinded manuscripts know or can fairly reliably guess authors' identities. But then blinding is worthless to

[6] We thank an anonymous reviewer for raising a worry along these lines.

[7] We thank Carlos Santana for discussion here.

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

try—and a waste of resources to implement. That suggests the norm is defective because it recommends blinding where it should not be done.

We do not think the objection undermines the norm. First of all, note that reviewers are not omniscient about authors' identities. In a review of studies across disciplines, blinding authors' identities from reviewers had a success rate ranging from 53 to 79% (Largent & Snodgrass, 2016, p. 87). But, just for the sake of argument, let us assume one of two states holds during a blind review of any manuscript: either (a) reviewers are 100% certain they know the author and they are right or (b) reviewers do not believe they know the author. If (b) obtains in a review, blinding encourages impartiality. But if (a) obtains, what are the consequences? One problem is that not hiding authors' identities from reviewers lets reviewers rubber-stamp their friends' papers and dish out unfair criticism of their rivals' work. We discuss that problem more in the next section, but notice a way around it: reviewers can be kept honest with accountability practices, such as publishing reviewers' names along with the article (a practice called open peer review). Blinding does not have to induce uncritical reviews. A further point is that even a small field might grow, and setting up review practices that allow newcomers to enter the discussion is valuable. Bias against outsiders seems likely when authors' names are known, so ensuring blinding in a small field is arguably good over the long run.

A third objection is that the norm of presumptive blinding carries with it a commitment to peer review, but that commitment needs a defense because peer review has critics.[8]

Our response to the objection is to deny that peer review has critics. Instead, there's only debate about *when* peer review should happen. For instance, some theorists insist that peer review should happen before publication, allowing the review process to certify research (Rowbottom, 2022); other theorists argue that review should happen after publication (Heesen & Bright, 2021). But notice the agreement on the significance of review. Both positions are at least compatible with review being blind—though for reasons we note in the final section, pre-publication review will be easier to blind successfully. It is also worth recording our suspicion: publishing research with no peer review will lead to informational chaos. At the end of the nineteenth century, a physiologist named Michael Foster, secretary of the Royal Society, argued that a centralized system of journals was necessary to improve the quality of published research (Csiszar, 2010, p. 416). "[A] great deal of work," wrote Foster, "is thrust upon the world which the world were much better without—work which is crude, unfinished, unmatured, a veritable sewage thrown into the pure stream of science, which has to be got rid of before the stream can again become free from impurity" (Foster, 1894, p. 728). To borrow the words of the political provocateur Steve Bannon, to publish research without any peer review is "to flood the zone with shit."

4. What Are the Downsides of Blind Review?

We have described the norm of presumptive blinding: when some information is determined to be task-irrelevant, it should be blinded, unless blinding is determined to have a significant downside. Here are four responses to the norm. First, endorse the norm and claim that blinding has no significant downside. We ourselves favor that response. Second, you might endorse the norm but insist that blinding does in fact have a significant downside, meaning you deny that known task-irrelevant information should always be blinded in review. Third, you might reject the norm by arguing it is false. A fourth response is to reject the norm by arguing that it has not yet been supported with sufficient reasons.

For now, we want to argue against the second response—that the norm is correct but blinding actually has a significant downside, meaning that blinding task-irrelevant information is *not* required.

[8] We thank Haixin Dang for discussion here.

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

Plausibly, blinding conflicts with many scientists' practices and values. As a result, some observers will view blinding as carrying a significant downside—and it will seem either that scientists' practices and values need revision or that blinding is a mistake. But then presumptive blinding is no longer in effect: the presumption is overridden, canceled, or defeated by countervailing considerations. We will discuss possible downsides involving editorial management, lack of accountability for editors and reviewers, and reduced author diversity in science. For each of those three concerns, the claim at issue will be that blinding introduces downsides that override whatever benefits it brings.

Let's start with editorial downsides potentially introduced by blinding. The thought here is that blinding ties the hands of editors and reviewers, preventing them from addressing authorial problems such as conflicts of interest and self-plagiarism. As one commentator noted, those problems "are more difficult to pick up in double-blind peer review because […] the reviewer does not know [the] author's identity" and "they cannot tell, for example, if the author is receiving funding" that constitutes a conflict of interest (Jarvis, 2017, p. 304). We believe that blind review can be fully consistent with editorial practices that aim to identify conflicts of interest and self-plagiarism. For instance, the editorial process could be designed so that before blind peer review, manuscripts are automatically checked for self-plagiarism against a database. (Such vigilance is better left to machines than humans.) Tiered review and registered reports can also help if checking for self-plagiarism and conflicts of interest happens after an initial blind review has concluded but before a final acceptance is issued.

A related possible downside is that editors plausibly have a duty, not only to maximize the quality of articles appearing in their journal, but to maintain—and if possible increase—the prestige of the journal. Accordingly, publishing work by well-known researchers is desirable. But when triple-blinding is the policy, editors can't do their best to improve the journal. And that seems to be a reason to limit blinding for editors.[9] Again, we believe that small tweaks could preserve blinding. Suppose the editorial workflow involves tiered review, where editors are temporarily blinded to authors' identities but later that information is unmasked. Editors can be permitted to peek at the names of authors before final verdicts are issued, and if there is competition for page space among equally good papers (ones deemed suitable for publication during blind review), "prestige" considerations can then enter the calculus. But insofar as editors might be too eager to publish work by high-status authors, or negatively disposed toward low-status authors, it seems useful to retain as much blinding as possible.

A second potential downside of blinding is reduced accountability for reviewers and handling editors. Dietmar Wolfram and colleagues express the concern as follows: "[b]linding of reviewer identities may allow reviewers to use their anonymity to deliver more critical reviews or to write reviews that lack rigor because authors and readers will not know who the reviewers are" (Wolfram et al., 2020, p. 1034). Blinding can become a sort of "Ring of Gyges," letting anonymous reviewers behave badly with impunity.

We concede—and know from experience as authors—that reviewers can be total schmucks. But let's not ignore an empirical question: Does less blinding actually encourage worse behavior than more blinding? For instance, does single-blind review, where reviewers know authors' identities, encourage reviewers to "go off" on authors who they disrespect or who would dare challenge their work? We find it plausible that stricter blinding policies can reduce unprofessional reviewer behavior.

Nevertheless, the benefits of both blinding and reforms to enhance reviewer accountability can be preserved through processes of tiered review and registered reports. One example is open peer review, where the identities of both reviewers and handling editors as well as reviewers' reports are published alongside articles (Wolfram et al., 2020). Reviewer identities can be blinded throughout

[9] We thank William Lycan for discussion here.

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

the review process, and those identities can be unmasked after the review process concludes. Blinding ensures impartiality in the review process, and unmasking reviewers' identities after publication ensures adequate accountability for reviewers and editors. Even in cases of rejection, unmasked reviewer reports can be shared with authors. Blinding can't change the fact that "Reviewer 2" is a natural kind. But it is not worth compromising the impartiality of peer review merely to identify Reviewer 2s—at least not when they can be identified effectively while preserving the benefits of blinding.

Finally, a further potential downside is reduced author diversity in publications. A common objection is that blinding disfavors marginalized authors. The marginalization might stem from gender, race, professional seniority, institutional prestige, geography, or some combination thereof. Titles from some recent articles announce the worry: "Blind Review Is Blind to Discrimination" and "Indigenous knowledges and scholarly publishing: The failure of the double-blind peer review." One possibility is that blinding leads to less diversity among authors; another is that it leaves the status quo intact, neither increasing nor decreasing author diversity; a third possibility is that blinding increases diversity among authors.

The matter can't be decided from the armchair. Unfortunately, data about the effects of blind review is complicated and inconclusive. In summarizing studies on blinding from many disciplines, Emily Largent and Richard Snodgrass note that, while "it is far from clear that blinding actually imparts fairness," the totality of evidence suggests that, at worst, blinding is ineffectual at increasing author diversity (2016, p. 92). Nevertheless, Largent and Snodgrass also note studies showing that blinding benefits marginalized groups, and they identify biases, such as prejudice against non-U.S. authors, that might be mitigated by double-blind review.

Obviously, blind review should not be expected to solve all of science's troubles with diversity and discrimination. But notice the difference between *blinding reducing diversity* and *blinding failing to increase diversity*. Consider those in order. First, at present, there is no evidence showing that blinding reduces diversity. Second, there is mixed evidence that blinding fails to increase some forms of diversity.

Why might blinding fail to increase author diversity? One explanation is that blinding fails to be up to the task—it cannot eliminate biases that work against marginalized authors. Take, for example, some research investigating "gender gaps" in grant proposals, which found that a gap in Gates Foundation funding was explained by differences in how male and female applicants used language (Kolev et al., 2020). That study might suggest that gender biases are sometimes not fully mitigated by blinding efforts.

There is an alternative explanation, however: the grant applications were not blinded enough. A case where *blinding fails to be implemented effectively* is critically different from a case where *blinding is incapable of reducing a certain bias*. Subtle cues in rhetoric and style are just one way that blinding can be compromised. In fact, there is good evidence showing that efforts to blind often fall short, such as when a reviewer guesses correctly an author's identity or an editor doesn't fully mask authorial information (Largent & Snodgrass, 2016, pp. 87–89). If the Gates Foundation grant applications were not blinded enough, that leaves open the possibility that stricter blinding procedures might better combat biases against marginalized authors. For instance, journal submissions could involve an initial stage of review with a bare-bones report about the findings, in order to minimize the influence of certain language use on evaluators' judgments.

While we think that implementing stricter blinding policies could result in greater author diversity, we believe it is possible that blinding simply will not eliminate certain biases in peer review against marginalized groups. Suppose that is in fact the case. What follows? Should we give up on presumptive blinding? No. Calls to eliminate blinding for the sake of diversifying authorship appear wrongheaded for a simple reason. Doing so would open the floodgates for many other biases that blinding is known to help mitigate. As we have already noted, blinding is effective at reducing status biases. And it is also clear from the available data that blinding does not amplify biases against

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

marginalized groups. Thus, even if blinding does not eliminate biases against diverse authors, it seems to have a clear net benefit when it comes to reducing bias.

5. Blind Review in an Interconnected World

In this essay, we have defended the norm of presumptive blinding. When some information is determined to be task-irrelevant, it should be blinded, unless blinding is determined to have a significant downside. We don't find any significant downsides that can't be resolved with relatively simple tweaks to current review practices. We do not say there *couldn't* be compelling downsides, but we just have not seen what those could be.

As we conclude, we want to suggest that blinding is a practice supported by a putative norm of science—what we call the "no free pass" norm. Consider Robert K. Merton's proposal that scientific claims are accepted or rejected wholly on the basis of conformity "with observation and with previously confirmed knowledge" (Merton, 1973/1942, p. 270). A corollary, Merton thought, is that "[t]he acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of their protagonist; his race, nationality, religion, class, and personal qualities are as such irrelevant" (Merton, 1973/1942, p. 270). Blinding facilitates science by forcing evaluators to ignore identity features of the "protagonist" and, in a tiered or sequential review, the specific conclusions they take their data to support. According to the no-free-pass norm, *scientific work should be evaluated independently of the scientist's experience, prestige, or ideological commitments.* The presumptive-blinding norm helps scientists follow the no-free-pass norm by preventing task-irrelevant information from entering evaluators' headspace. We noted above that blinding is a practice of "brutal humility," imposing on evaluators a more impartial way to evaluate peers' ideas.

But our reflection on blinding has made us unsure how easy it is to follow both the presumptive-blinding norm and the no-free-pass norm. In the internet age, researchers often post unpublished manuscripts and data on preprint servers and social media platforms. Similar practices predate internet technologies, to be sure—researchers present at workshops and circulate manuscripts in peer networks. Sharing unpublished work allows scientists to receive feedback, speed up dissemination of their ideas, and pump up their careers.

Imagine a preprint goes viral on a social media platform. Later on, the authors submit their work to a journal. Many of the invited reviewers will have already seen—and indeed read—the submitted manuscript and know the authors' identities. That means blinding is not successfully implemented —through no fault of the editor or journal. The preprint server and social media platform broadcasts the research but also the authors' identities. Scientists have a personal stake in publishing their work and, in times of crisis, rapidly disseminating their knowledge. All of this raises a question: How can we retain the benefits of pre-publication publicity that accrue to scientists and broader communities while at the same time preserving the practices of blind review? That is a difficult question and one we have not seen addressed head-on.

Now imagine at the snap of our fingers we could make every major journal in every scientific field adopt double- or triple-blind review. Would that be a big win for the presumptive-blinding and no-free-pass norms? We suspect it wouldn't be quite as good as it initially seems. That's because the publicity of pre-publication science is pervasive enough to compromise many instances of review that would have been effectively blinded otherwise. Blinding pulls one way, publicity another. How much are scientists willing to compromise on impartiality to accommodate the social and reputational functions that publicity serves?

Importantly, we do not view the presumptive-blinding and no-free-pass norms as standing in fundamental tension with calls for open science. For example, as we noted earlier, the practice of open reviewing is compatible with blinding whenever authorial and reviewer information is disclosed *after* a blinded review process concludes. Registered reports are another example: these incorporate aspects of blinding with various open science practices, including open data and open registration. Initial evidence suggests that registered reports yield higher quality publications than

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

the traditional reviewing model (Soderberg et al., 2021), and there are surely other ways in which blinding and open science practices can complement each other en route to improving science (Waltman et al., 2023). Nevertheless, some practices, such as preprinting, are not as easy to reconcile with the norms we've discussed in this essay.

Our suggestion is that the right norms without buy-in won't improve science. If the presumptive-blinding and no-free-pass norms are indeed correct, as we believe they are, then scientists must have hard conversations about whether their cultural practices and institutions respect all of the norms that science should embrace.[10]

References

Aldhous, P. (2014). Scientific publishing: The inside track. *Nature*, 510, 330–332. https://doi.org/10.1038/510330a

Ballantyne, N. (2015). Debunking biased thinkers (including ourselves). *Journal of the American Philosophical Association*, 1(1), 141–162. https://doi.org/10.1017/apa.2014.17

Ballantyne, N. (2022). Tragic flaws. *Journal of the American Philosophical Association*, 8(1), 20–40. https://doi.org/10.1017/apa.2020.39

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures*, 12(1), 11–45. https://doi.org/10.1080/1045722022000003435

Button, K. S., Bal, L., Clark, A., & Shipley, T. (2016). Preventing the ends from justifying the means: Withholding results to address publication bias in peer-review. *BMC Psychology*, 4(1), 1–7. https://doi.org/10.1186/s40359-016-0167-7

Celniker, J. B., & Ditto, P. H. (in press). Of preferences and priors: Motivated reasoning in partisans' evaluations of scientific evidence. *Journal of Personality and Social Psychology*.

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour*, 6(1), 29–42.

Cheek, N. N., & Pronin, E. (2022). I'm right, you're biased: How we understand ourselves and others. In N. Ballantyne & D. Dunning (Eds.), *Reason, bias, and inquiry: The crossroads of epistemology and psychology* (online edition). Oxford Academic. https://doi.org/10.1093/oso/9780197636916.003.0003

Clark, C. J., Isch, C., Everett, J. A. C., and Azim Shariff. (2023). Even when ideologies align, people distrust politicized institutions. *PsyArXiv* (April 14). https://doi.org/10.31234/osf.io/sfubr

Csiszar, A. (2010). Seriality and the search for order: Scientific print and its problems during the late nineteenth century. *History of Science*, 48(3–4), 399–434. https://doi.org/10.1177/007327531004800306

Csiszar, A. (2016). Peer review: Troubled from the start. *Nature*, 532, 306–308. https://doi.org/10.1038/532306a

Csiszar, A. (2018). *The scientific journal: Authorship and the politics of knowledge in the nineteenth century*. Harvard University Press.

Cuddy, A., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, 40, 61–149.

Cusimano, C., & Lombrozo, T. (2023). People recognize and condone their own morally motivated reasoning. *Cognition*, 234, 105379.

Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D. E., Saks, M., & Risinger, M. (2015). Context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, 60(4), 1111–1112.

Foster, M. (1894). An address on the organisation of science: Delivered before a general meeting of the XIth International Medical Congress, held in Rome, 1894. *BMJ*, 1, 727–728.

Fox, C. W., Meyer, J., & Aimé, E. (2023). Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Functional Ecology*, 37, 1144–1157. https://doi.org/10.1111/1365-2435.14259

Heesen, R., & Bright, L. K. (2021). Is peer review a good idea? *British Journal for the Philosophy of Science*, 72(3), 635–663.

[10] For comments and conversations, we are grateful to Matthew Altman-Suchocki, Alex Arnold, Andrew Bailey, David Christensen, Alex Csiszar, Adrien Fillon, William Lycan, and Benjamin Wilson. We also thank students in NB's seminar during spring 2024, and members of the Philosophy, Cognition, and Culture Lab at Arizona State University, especially Marcello Di Bello, Peter Kung, Shyam Nair, and Keith Swisher. An earlier version of this material was presented during a workshop on the Epistemology of Science Communication at the University of North Carolina in September 2023, and we thank participants at the workshop for their questions. Special thanks to our commentator Corey Cusiamano as well as Haixin Dang and Carole Lee. Finally, two anonymous reviewers provided valuable input on a previous version. (Although NB is a co-editor of the *CJP* special issue in which this essay is published, he recused himself during the full review process and thus did not select the reviewers or otherwise learn of their identities.) Finally, our writing was supported by a generous grant from the John Templeton Foundation (grant 62824).

https://doi.org/10.1017/can.2024.25 Published online by Cambridge University Press

Huber, J., Inoua, S, Kerschbamer, R, König-Kersting, C, Palan, S, & Smith, VL. (2022). Nobel and novice: Author prominence affects peer review. *Proceedings of the National Academy of Sciences*, 119(41), 36194633. https://doi.org/10.1073/pnas.2205779119

Jarvis, S. (2017). The perils of peer review. *Veterinary Record*, 181, 304. https://doi.org/10.1136/vr.j4362

Kelly, T. (2023). *Bias: A philosophical study*. Oxford University Press.

Kolev, J., Fuentes-Medel, Y, & Murray, F. (2020). Gender differences in scientific communication and their impact on grant funding decisions. *AEA Papers and Proceedings*, 110, 245–49. https://doi.org/10.1257/pandp.20201043

La Rochefoucauld. (1678/1959). *Maxims* (Translated with an introduction by L. Tancock). Penguin.

Largent, E. A., & Snodgrass, R. T. (2016). Blind peer review by academic journals. In A. S. Kesselheim & C. T. Robertson (Eds.), *Blinding as a solution to bias: Strengthening biomedical science, forensic science, and law* (pp. 75–95). Academic Press.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64, 2–17. https://doi.org/10.1002/asi.22784

Merton, R. K. (1973) [1942]. The normative structure of science. In R. K. Merton (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 267–278). University of Chicago Press.

Pleskac, T. J., Kyung, E., Chapman, G. B., & Urminsky, O. (2023). Blinded versus unblinded review: A field study comparing the equity of peer review. https://doi.org/10.31234/osf.io/q2tkw

Rowbottom, D. P. (2022). Peer review may not be such a bad idea: Response to Heesen and Bright. *British Journal for the Philosophy of Science*, 73(4), 927–940.

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, 5(8), 990–997.

Thompson, W. C. (2016). Determining the proper evidentiary basis for an expert opinion: What do experts need to know and when do they know too much? In A. S. Kesselheim & C. T. Robertson (Eds.), *Blinding as a solution to bias: Strengthening biomedical science, forensic science, and law* (pp. 133–150). Academic Press.

Tomkins, A., Zhang, M, & Heavlin, WD. (2017). Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of the Sciences*, 114(48), 12708–12713. https://doi.org/10.1073/pnas.1707323114

Tuvel, R. (2017) In defense of transracialism. *Hypatia*, 32(2), 263–278. https://doi.org/10.1111/hypa.12327

Uhlmann, E. L. (2011). Post hoc rationalism in science. *Behavioral and Brain Sciences*, 34(4), 214. https://doi.org/10.1017/S0140525X11000410

Waltman, L., Kaltenbrunner, W., Pinfield, S., & Woods, H. B. (2023). How to improve scientific peer review: Four schools of thought. *Learned Publishing*, 36(3), 334–347.

Weller, A. C. (2001). *Editorial peer review: Its strengths and weaknesses*. Medford, NJ: Information Today.

Wolfram, D., Wang, P., Hembree, A., & Park, H. (2020). Open peer review: Promoting transparency in open science. *Scientometrics*, 125(2), 1033–1051.

Cite this article: Ballantyne, N., & Celniker, J. 2024. No Peeking: Peer Review and Presumptive Blinding. *Canadian Journal of Philosophy*: 1–14, doi:10.1017/can.2024.25