

## **Publishing, Belief, and Self-Trust**

Alexandra Plakias  
Hamilton College  
[aplakias@hamilton.edu](mailto:aplakias@hamilton.edu)

Submitted: 21 July 2022

Accepted: 25 Sept 2022

*Abstract* This paper offers a defense of ‘publishing without belief’ (PWB)—the view that authors are not required to believe what they publish. I address objections to the view ranging from outright denial and advocacy of a belief norm for publication, to a modified version that allows for some cases of PWB but not others. I reject these modifications. In doing so, I offer both an alternative story about the motivations for PWB and a diagnosis of the disagreement over its permissibility. The original defense focused on consequentialist reasons for allowing PWB, offering mostly defensive arguments against potential criticisms. But I argue that once we shift our focus to the reasons why authors might be prone to PWB, we see a difference in two types of motivation: whereas I imagine PWB as arising from underconfident agents, critics point to cagey or nefarious authorial practices, or authors’ failure to clarify their own degrees of belief. Underlying the debate over norms of philosophical publishing, we find two different conceptions of philosophy itself.

*Keywords:* publishing ethics, self-trust, confidence, metaphilosophy, social epistemology, virtue epistemology, epistemic injustice

### **1. Introduction**

Becoming a professional academic philosopher involves producing research, usually in the form of academic papers. We train our graduate students in the mechanics of doing so. But we spend less time on the question of how to arrive at the ideas for said papers, and how one ought to relate to those ideas. Professional philosophers come to be identified with their views, such that we assume that the views someone defends in print are the views she believes. Is such an assumption warranted? Empirically, the answer is probably yes. As a normative matter, things are more complicated.

In an earlier paper (Plakias 2019), I defend a view called, ‘publishing without belief’ (PWB), according to which there is no norm requiring that authors believe what they publish. Responses have ranged from outright denial of this view and advocacy of a belief norm for publication (Buckwalter forthcoming), to modifying my original claim and allowing for some cases of PWB but not others (Fleisher 2020, Sarihan 2022).

Here, I defend and develop original position, introducing further arguments against a belief norm for publication and rejecting the proposed modifications of the view. In doing so, I offer both an alternative story about the motivations for PWB and a diagnosis of my disagreement with critics. My original argument focused on consequentialist reasons for allowing PWB, offering mostly defensive arguments against potential criticisms. But once we shift our focus to reasons why authors might be prone to PWB, an underlying difference becomes clear: whereas my (2019) imagines PWB as arising from underconfident agents, critics point to cagey or nefarious authorial practices, or authors’ failure to clarify their own degrees of belief.

My earlier argument envisioned an author who is not confident enough in her arguments to believe them, leading critics to suggest that the hypothetical author modify her claim. But the strongest defense of PWB—the defense I offer here—views it as arising, not out of (lack of) confidence understood as a purely epistemic relation of an author to her arguments, but out of a (lack of) confidence understood as self-trust (Jones 2012). Once we reconceive the central issue this way, it becomes clear that parties to the debate are talking past one another, with one side (me) addressing arguments to a non-ideal conception of philosophy, and the other side (critics of PWB) addressing arguments to an ideal conception.

As before, the discussion here is restricted to philosophical publishing. That’s not the case with all of the arguments I discuss (eg Buckwalter and Habgood-Coote, who explicitly include examples involving scientific publication). Other authors have focused exclusively on the role of belief in scientific publishing (Dang & Bright 2021). My reasons for focusing on philosophy are related to some of the arguments I’ll defend later in the paper: namely, that we should approach this issue with specific aspects of actual philosophical practice in mind, and evaluate norms

governing publication with an eye to the features that affect authors' confidence in their work—and these may be particular to philosophy. If the arguments here generalize to other disciplines, that's fine too.

The paper begins by reviewing the original argument for PWB and the responses. I show how reliance on cases and the intuitions they generate risks biasing and confusing our intuitions against the practice. I then survey the main objections to PWB: it withholds or misrepresents evidence; it erodes trust; it is unnecessary because authors ought to qualify their claims so that they accurately represent belief. After explaining why each of these criticisms fails, I take up the question of confidence. I've argued that PWB would disadvantage the more epistemically humble or virtuous among us; critics argue that we can simply advance our arguments more modestly. I suggest that there are two senses of confidence at work here, and once we distinguish them, we can see what's really at issue in the debate: the two sides are evaluating two different pictures of philosophy, one idealized, and one less-than-idealized. On a nonideal epistemology of publishing, I argue that my picture is, while not flattering, correct.

## **2. Publishing Without Belief: Problematic Authors**

Authors may advance claims and arguments in their published work without believing them, and are not therefore criticizable for doing so. The arguments for this claim presented in my earlier paper are largely defensive: rather than arguing that PWB is good in itself, I suggest that there are undesirable costs to endorsing a norm that requires belief, and few positive reasons to do so. For brevity's sake, I'll refer to the position I argue against—the idea that there is in fact a norm that requires authors to believe what they publish—as the Belief Norm of Publishing (BNP). Since I reject the norm, I won't spend too much time on how we ought to specify it; I leave it open whether BNP is entailed by some stronger norm which itself best represents the norms governing publication (for example, if the norm of publishing were to be a knowledge norm, this would require belief but be much stronger), or sits alongside other norms which also govern the act of publication (as argued in Levy forthcoming). For example, Buckwalter argues directly for a belief norm; Habgood-Cote (forthcoming) argues for a 'sincerity norm' requiring belief via the claim that publishing is a species of assertion.

I've argued (2019) that sanctioning PWB would disadvantage the more epistemically virtuous members of the philosophical community, as well as those inclined towards conciliationism (not necessarily the same group). That's because these individuals might be more inclined to see philosophical disagreement as a reason for doubting their view, whereas steadfast individuals or the epistemically arrogant (again, not necessarily the same group) will be less troubled by it.

In addition to this consequentialist argument, the earlier paper presents a variety of cases intended to both characterize instances in which we might find PWB and elicit intuitions about its permissibility. Critics have responded with their own cases, designed to elicit opposing intuitions. But intuitions are a poor guide here, regardless of where they happen to fall. That's because the cases offered by each side involve multiple types or instances of epistemic malfeasance, making it difficult to identify the source of our disapproval: is it the author's lack of belief, specifically, or some related but distinct factor that generates our reaction? While this complication limits the probative value of our intuitions, it also highlights an important feature of the debate: a belief norm for publishing isn't the only basis for criticizing bad epistemic behavior by authors in these cases. There are other grounds on which to condemn them.

Two of my original three cases involve authors who have doubts about their argument or the evidence for it, and I'll say more about these below. The third case is more problematic, and involves an author who actively hoaxes a journal. I ask, "Do any of these philosophers merit criticism? Have they erred, either as epistemic agents generally or as philosophers specifically?" and answer in the negative (2019: 639).

Critics have not shared my diagnosis. Most discussions of the norms of publishing agree that authors who submit 'hoax' papers violate a norm of some sort, and belief or sincerity are obvious candidates. I agree that the case is problematic, and the worry that PWB opens the door to hoaxing or 'trolling' journals is a serious one. As Buckwalter (forthcoming) writes, "there is probably no quicker way to undermine public trust in research than to normalize the publishing of papers hoaxing it."

In this case, our intuitions are likely distorted by our moral judgment of the hoaxer. Let's stipulate that hoaxing journals is morally bad, partly for the reasons outlined above: it's a way of lying, it wastes resources, etc. We want to condemn the hoaxer in this case, and the obvious way to do so—given the context—is by condemning him qua author. But this isn't the only way to criticize the agents presented in these cases. It's too quick to claim that any of the agents in the cases, "have erred... as epistemic agents generally" and "as philosophers specifically," (2019: 639). A better strategy would be to concede error and locate it not in the act of publication, but in something prior: a false belief, sloppiness, or some other epistemic fault. We already recognize a variety of epistemic vices; we criticize philosophers whose work evinces them.

The defender of PWB can thus agree that hoaxing is bad, while disagreeing over the reason why: the problem is not PWB per se, but a more general epistemic vice or violation. Hoaxers seem to be fundamentally misrepresenting their work-- not because they don't believe it, but because they don't take it to be academic research to begin with, but rather a kind of parody or reductio of such. It's fairly (if not entirely) uncontroversial that inquiry aims at truth; given that hoaxers are not aiming at truth, they're not engaged in inquiry, and thus they're not actually doing research. Instead, it's a kind of performance. But there's a significant difference between fundamentally misrepresenting one's enterprise and not divulging doubts; the hoaxer is doing the former, while the authors in the other cases are doing the latter.

Another instructive case is Fleisher's (2020) 'poorly read ethicist' Nelson, who "writes an overview article... for an online encyclopedia," but has "negligently only read a small portion of the papers he refers to," and as a result believes few of the claims he publishes about these papers and their arguments. Describing this and two other cases, Fleisher writes, "each of these cases involves publishing without belief... each also involves epistemically and philosophically inappropriate behavior... [the authors] have erred because they have published claims they do not believe." Crucially, "they do not believe the claims precisely because doing so would be unjustified." (239-40)

Nelson is epistemically problematic. But wherein lies the problem? Like Brad the hoaxer, Nelson's epistemic sins predate the act of publication: in failing to adequately research his

article, Nelson shirks his assignment. In fact, I think there are at least two alternative grounds for criticizing Nelson: first, in failing to adequately research his subject, Nelson is failing in his role as an encyclopedia entry author. (One might wonder whether the fact that he's writing a reference work subjects him to special norms—I'll say more about this below.) Second, Nelson seems to exhibit more general epistemic failures: not only is he "negligent," but he seems careless; insouciant; lazy. PWB *can* result from epistemic insouciance (Plakias 2019: 643), but need not. Importantly, when it does, we have grounds to criticize the authors not in their capacity as authors, but as epistemic agents more generally. So, our intuitions in Nelson's case are elicited in part by his epistemic character (as revealed in his shoddy research), but we might mistakenly attribute them to the act of publication itself.

Nelson's case holds two lessons: first, there are epistemic norms that philosophers ought to follow in publishing, or in their inquiry more generally, and to the extent that they violate these, we can still condemn them. A philosopher's lack of belief might stem from his knowledge that he has in fact violated some such norm, in which case we judge them for *that* violation, not for a violation of BNP. Fleisher considers something like this possibility, acknowledging that, "One might worry that... in each of my cases it is the author's lack of epistemic standing... that explains why their publishing is impermissible." (241) However, he maintains that Nelson's lack of belief is itself problematic, because if Nelson adequately performs his duties, "the evidence is sufficient to justify his attempt to get others to believe the claims," and "this should lead him to form a justified belief on the matter." But while Nelson's lack of belief is a failure of rationality, it is not a failure of his epistemic duty as an author—if we expect authors to form beliefs based on evidence, this is in virtue of a larger epistemic obligation, not something specific to the institution of publishing.

The second lesson is that our intuitions can mislead us. Our thought that Nelson has violated *an* epistemic norm can become the thought that, since the violation happened in the course of publishing, the norm violation must originate in the act of publication itself. BNP is a natural magnet for these intuitions, because it's ambiguous between a true descriptive claim and a false normative claim. That is, it's descriptively true that most authors *do* believe what they publish, but it's false that they commit a norm violation if they don't.

One might argue that such an author violates our expectations. We expect authors to believe their published work, but this is an empirical expectation-- authors usually *do* believe their work. We also expect authors to carry out certain duties and perform certain behaviors related to their work: we expect them to continue to defend arguments after they're published, give talks on the work they've published and stand behind the arguments, and even develop expertise in the area of those arguments.<sup>1</sup> So, an author who publishes a paper defending moral realism and then immediately pivots to attacking it violates empirical expectations of consistency, and may violate certain professional expectations relating to giving talks, defending published work, and so on. But again, we can criticize her for that without criticizing her belief (or lack thereof) itself. Criticizing someone for failing to perform a role is distinct from criticizing their attitude towards that role.

However, it's true that believing in one's arguments makes investing the time required to research, write, and publish much easier. That doesn't mean it's impossible to perform the role without the attitude, but it means that a norm like BNP isn't the only way to exert pressure on authors to believe their work—the practical commitments that go along with publication will also have an effect.

Indeed, one might worry that they exert *too* much of an effect. I'm focusing on reasons an author might fail to believe her work, but the reasons why she believes it might themselves be epistemically problematic. The phenomenon of 'advocacy bias' demonstrates that at least in some cases, the act of defending a position shifts our belief in favor of that position—when we're tasked with defending a view, we become more convinced of its truth, and less adept at identifying flaws. As Melnikoff and Strohming (2020) point out (while offering empirical evidence in favor of the effect) we find discussion of the bias going back at least as far as Plato's *Gorgias*. Advocacy bias demonstrates that the correlation between belief and publication can take one of two causal paths, and it's not always an epistemically flattering relationship: sometimes, the fact that we believe a view isn't the motivation for publishing a defense of it, but

---

<sup>1</sup> For a detailed discussion of authorial obligations and roles, see Habgood-Coote (forthcoming).

a consequence of writing about it. Belief can track the evidence in favor of a position, but it can also lead us to overlook evidence against it.

The lesson: it matters why the author does or doesn't believe their work. If an author doesn't believe his work because of some prior epistemic malfeasance on his part, then our judgment of wrongdoing should be directed at that, and not at the act of publication itself. Neither my original defense of PWB nor the one I offer here is intended to show that there are *no* norms of publication; that would be an extreme position indeed. 'Don't publish things you know to be false' is a plausible norm; so is, 'don't falsely implicate you've read things.' (For additional discussion of the norms of publishing, see Levy forthcoming.) These norms, along with general epistemic principles and virtues of diligence, offer us ample grounds for critiquing authors.

### **3. Belief and Evidence**

Another objection to PWB involves the relationship between belief and evidence. Here again critics seem to be targeting imagined authors' lack of evidence or lack of transparency surrounding their evidence, rather than the act of publication itself.

For example, Buckwalter draws on the involuntariness of belief to support BWP: since we don't control our beliefs, and beliefs are formed "respective of truth-conducive considerations," the fact that we believe something is *prima facie* evidence for truth: "if belief aims at truth, and belief is the norm of academic publishing, then it follows that published claims will be more likely informed by truth-conducive considerations than those that are published but not believed... the fact that one believes a proposition is always at least some evidence for its truth." Here, the thought is that belief is an involuntary response to evidence for truth, so requiring belief "increases the likelihood that fewer false claims are made." (5) It's true that requiring a BNP makes it more likely that authors will fulfill their duties, because as we saw above, belief tends to correlate with the behaviors we expect of authors (defending their claims, pursuing them further, being accountable for them). But we should not let this mislead us into thinking that it's belief itself, rather than the corresponding behaviors and epistemic diligence, that serve as the norm for publishing.



The motivations that come with belief aren't always so laudable, either. The literature on cognitive biases suggests that belief can be involuntary in response to non-truth conducive factors. One example (discussed above) is the advocacy bias; wishful thinking is another. And the type of evidence on which a claim is also relevant: perhaps more than other fields, philosophy relies on intuitions as evidence, and these too may rely on factors outside our control.<sup>2</sup>

The discussion thus far has focused on whether PWB indicates shortcomings in the author; other arguments focus on the ways PWB might shortchange readers. Fleisher (2020) and Sarihan (2022) argue that PWB is criticizable insofar as it either withholds or misrepresents evidence to readers. After all, Sarihan notes, there must be a reason why authors don't believe their arguments. Either this reason has bearing on the evidence for (or against) the argument, or it's trivial. In the former case, the author is effectively withholding philosophical evidence, and this crosses a line: "PWB is impermissible when it involves withholding substantive reasons for disbelief." (2)

But what relationship do the author's reasons have to the reader's? We should all agree with that, 'famous philosopher Y believes p' is not itself good evidence for p (Plakias 2019: 644). We might think it's evidence *that there is reason to believe* p—but that reason isn't the philosopher's belief itself; the belief is a proxy for some other evidence (Kelly (2005) makes a similar point in defense of the steadfast response to peer disagreement). In the case of publication, we are expected to offer that evidence directly. This is why we review papers blindly: we want to know the author's evidence, not the author's identity. If I review a paper, recommend it for publication, and subsequently discover that the author has just published a paper arguing for exactly the opposite view, I should be surprised, but I should not change my verdict, nor second-guess it.

Fleisher (2020) takes a more fine-grained approach to the question of what role an author's claims play for a reader. He agrees that PWB is sometimes permissible, but argues that this depends on the type of claim an author is making. Fleisher distinguishes *advocacy role claims*,

---

<sup>2</sup> Unlike Plakias, Buckwalter doesn't restrict his claims to philosophy—he intends BNP to apply to science as well.

which “aim (or function) to promote productive debate or disagreement,” (242) from *evidential role claims*, which “aim or function to increase the common stock of evidence available to inquirers.” (243) An author need not intend his audience to believe advocacy claims; indeed, he might expect they will not, and will instead add their own counter-claims. On the other hand, evidential role claims aim to add to the “common stock of evidence,” and so must be not only well-justified, but made in a kind of epistemic good faith. Advocacy claims generate debate; they are usually *not* taken on trust by a reader. Evidential claims, on the other hand, are believed in virtue of the reader’s trust in the author. Thus, ERCs “require a higher epistemic standing, one that does include a belief requirement.”(237)

As the name suggests, ERCs aim to add to the set of claims we (the community of inquirers) accept. Thus, it would be impermissible for an author to publish an ERC without justification, or based on shoddy evidence. To make an ERC, one must be in a position to justifiably believe it; that justification then gets passed to readers. But if one *is* in a position to justifiably believe the claim, why wouldn’t one actually believe it? I’ll answer that question in a bit. For now, notice that this claim—that one needs to be in a position to justifiably believe the ERC—is different from requiring belief full-stop.

There’s also the question of whether the ERC/ARC distinction is tenable. Fleisher grants that most papers will contain both types of claims—few cases will be as clear-cut as Nelson’s aforementioned encyclopedia entry. Even within encyclopedia entries, what serves as an ERC for one reader may appear to another more expert reader as an ARC (for example, an expert might take issue with the way a certain philosopher’s view is classified, or with the author’s definition of moral realism). We could appeal to the author’s intention to settle the question, but since the distinction is partly motivated by the different function played by the two types of claims, authorial intention can come apart from actual function here, as Fleisher points out (242). A single claim can play either role, or both: what functions as an ERC for a more novice reader can be an ARC for a more expert philosopher working in the area. For example, the claim that a certain historical figure’s view is an example of moral realism might be an ERC for a student of metaethics, but an expert might find something to argue with in the claim. Furthermore, a claim’s

meaning might evolve over time: as canons of evidence shift, what was once controversial becomes common knowledge (and vice-versa).

That's not to say the distinction can't be made, or that there's no a difference between these two roles for philosophical claims. But it is a problem for Fleisher's proposed restriction of PWB. On Fleisher's view, PWB is permissible only when it comes to ARCs, and not for ERCs. But if a single claim can be read as playing either role, which norm should it be governed by? Do we read authors as responsible for believing particular claims in articles, or for their papers as a whole? One strategy would be to adopt the more stringent norm for the paper, given that parts of it are likely to contain ERCs and therefore be subject to it. Another would be to err on the side of permissivism, allowing PWB for both types of claims. As I've already indicated, I favor the second approach. Before explaining why, let me address one last objection to PWB: that it erodes trust.

#### **4. Trust**

The critics discussed here all express concerns about trust as a reason to be wary of PWB. Specifically, they argue that PWB would erode readers' trust in authors. Recall Sarihan's claim that PWB is wrong because it withholds evidence from readers. We might wonder why this withholding is wrong; one answer is that the author violates her readers' trust. This raises interesting questions about what we expect from other philosophers, and what we're entitled to expect. As a descriptive matter, I doubt many of us would be shocked to learn that our peers were working under "a mixture of time constraints and careerist motivations" (Sarihan 2022). One argument for BNP is that it reins these other influences in. Even if we reject Buckwalter's claim that belief is more likely to lead to true claims being published, we might think it is a useful way of "curating the research record," as Buckwalter puts it. Fleisher's argument for requiring authorial belief in ERCs is based on the idea that readers accept these these claims on trust, and that trust should be warranted. All of this raises the question: what role does trust play in philosophical argument and justification?

This is a big question, and I won't answer it here.<sup>3</sup> However, we can make some preliminary observations, focusing on publishing. Distinguish two ways of being trustworthy as a philosopher: you should believe what I believe, and you should believe the claims I present. These come apart. Someone might be insecure, or underconfident, or overly skeptical, and therefore her beliefs where her own work is concerned are an unreliable guide to truth. Someone might be unduly influenced by religion, or the philosophical position of their graduate advisor.<sup>4</sup> The many reasons someone's *belief* might fail to be a trustworthy guide to the truth come apart from the reasons her *arguments* might so fail. Indeed, a kind of epistemic caution and/or conscientiousness might make an agent simultaneously trustworthy as a guide to evidence and untrustworthy as a guide to the value of her own beliefs. If that sounds pathological, maybe it is: underconfidence among philosophers is a real phenomenon, and we ought to acknowledge and accommodate it rather than selecting against it via our publication norms.<sup>5</sup> (I say more about this in the final section of the paper, because I think there is a stronger/more persuasive version of the point.) Epistemic humility is something we as a discipline ought to promote, and sanctioning PWB risks discouraging it (for more on the ways philosophy might promote epistemic humility, see Kidd 2016).

Sanctioning PWB doesn't necessarily stop a proliferation of falsehood, but it does risk a proliferation of dogmatism and overconfidence. It rewards authors who have greater (and perhaps unearned) confidence in their own opinions, and a willingness to treat their belief as sufficient for entering evidence into the common ground. A further worry is that the willingness to see one's own belief as epistemically sufficient for evidence requires a degree of confidence that, in philosophy at least, is not evenly distributed amongst members of the profession. It tends to be rarer in more junior members, and its distribution may well be subject to other sorts of bias (see Kidd 2016 for further discussion and an argument to this effect). Returning to the ERC/ARC distinction, the extent to which claims are treated as evidence, or as an invitation to

---

<sup>3</sup> For extended discussion, see Levy 2022 and Levy (forthcoming).

<sup>4</sup> The canonical example here is Lackey's (1999: 477) 'Creationist Teacher,' who teaches her class about evolution despite not believing it herself. Lackey argues that she can nonetheless transmit knowledge to her class. This is a controversial case, and my intention here is not to focus on testimony per se, because like Plakias, I want to restrict the discussion specifically to publishing. My point is simply that we can trust someone by accepting their statements, and that this can be different from trusting them by believing what they believe.

<sup>5</sup> This was one of the motivations for my 2019 defense of PWB, though perhaps the point there is implied rather than stated outright. See also Plakias 2020.

debate them might be, too. If the question an author must ask herself is, ‘will people treat my claims as evidence, or as grounds for debate,’ then unfortunately, the author’s identity is not irrelevant: a junior woman’s claim that historical text X says Y may be more likely to be interpreted as an ARC (vs an ERC) than her senior male colleague’s claim to the same effect.

Perhaps this is not entirely unfortunate. Buckwalter suggests that in certain cases, an author’s lived experience is relevant to the assessment of her argument, and that this is yet further reason to require a BNP—since it would require “extensive academic detective work” to discover that researchers don’t “do, value, or experience what they claim to,” we’re forced to accept authors’ claims about their experiences and values. The role of lived experience as evidence in philosophy is a big issue, well beyond the scope of this paper. But the general question of the author’s role in justifying belief—that is, the role of the person, rather than the argument itself—is relevant to the debate here, and worth pausing to consider.

The question is whether we do accept the evidence presented in papers based on our trust in their authors. I doubt that we do— papers are initially granted publication based in large part on blind review, which indicates that we can judge the merits of an argument or claim independently of knowing anything about an author. But let’s grant the claim for the sake of argument. Even then, our trust is not based on the author’s *belief*. Fleisher writes, “when an author publishes a claim with the aim of having their audience believe it on trust, they are required to (justifiably) believe that claim.” At first glance, this seems to be precisely what I want to dispute. But notice the specified aim here: Fleisher is talking about a case where authors intend their readers to accept claims *based on trust*. After claiming that ERCs should be governed by an epistemic standard that requires an author to believe her claims, Fleisher writes, “Failure to meet this standard will lead to a proliferation of falsehood.” (7) To be sure, his cases involve authors publishing the kind of work no one should be happy to have around. But this is not because they don’t abide by a belief norm, it’s because they don’t abide by more basic norms about due diligence and using reliable sources. That’s what leads to the proliferation of falsehood (that, and lax refereeing). To the extent they violate our trust, it’s not because we trusted that they *believe* their claims; we trusted them to read the relevant literature, use reliable data-gathering surveys, etc.

Nor is our trust based on the *author's* belief. Construing trust as a relation between author and reader is too narrow; the reader's trust is placed in a system that includes referees, editors, and authors. The discussion of trust often focuses on the perspective of the reader, but if there is trust involved in this process, authors also rely on it: they trust the referees and editors to assess their arguments fairly and to request only those changes that will improve the article (and not, say, changes that will bring greater attention to their own work). And for reasons I'll discuss below, authors' trust in referees might go beyond trusting them to fairly assess their work; authors may use referees as a kind of proxy for belief, relying on referees' judgments that an argument is publishable in lieu of their own assessment of quality.

## **5. Confidence and Self-Trust**

This brings us back to questions of confidence and self-confidence. If we require authors to believe what they publish, authors with more confidence in their conclusions enjoy an advantage over less-confident authors. I've expressed worries that this disadvantages the underconfident philosopher; others (Buckwalter, Sarihan) have suggested that this worry can be avoided if less-confident authors qualify their claims appropriately. Thus, instead of claiming to have 'defended non-naturalism' an author might write that they have identified a possible response to an objection; instead of arguing that empirical evidence supports their view, an author might acknowledge doubts about the evidence but frame the argument as conditional: *if* the evidence is as it seems, then certain philosophical conclusions follow.

There are two problems with this line. The first is that the qualified claims are unlikely to succeed or receive as much attention. Buckwalter argues, "careful qualification and hedging are common features of academic writing." This is an empirical question, but I doubt they're that common. Consider one of his examples, involving the hypothesis that we live in a simulation. Buckwalter considers three claims a researcher might publish: the claim that we do live in a simulation; that the evidence strongly supports that we live in a simulation; that the evidence is about 50-50. Now, as a matter of fact, the last claim has been published, which might seem surprising. But it's *less* surprising when we consider that the first two claims have also been published. It's possible, of course, that the authors of the first two claims believed what they

published—I have no way of knowing, and neither does anyone except the authors themselves. What matters here is that insofar as the carefully hedged claim succeeds in generating attention, it does so against the background of the stronger, more sweeping claim.

The bigger problem is that the reply misidentifies the kind of confidence at issue. There are two senses of confidence at issue in the debate over PWB, and ambiguity about which one we're interested in accounts for some of the disagreement between its defenders and opponents.

There's an epistemic sense of confidence, in which 'confidence' refers to the degree of belief a philosopher has in her claims, her work, and her arguments; this is one way of assessing her as under- or over-confident, qua epistemic agent. And there's another, non-epistemic sense, in which confidence refers to the relationship an agent has to her own judgment.

Discussions of PWB so far (see especially Plakias 2019 and Buckwalter forthcoming, especially his discussion of qualifying claims) seem to be using confidence in an epistemic sense, where confidence is a matter of apportioning or calibrating our degrees of belief. On this view, confidence is an epistemic attitude; a relationship between an agent and her belief. The more she thinks a belief is true, the more confident she is in it. If I'm overconfident in my belief that my paper will be accepted, I rate it more likely to be true than the evidence warrants; if I'm underconfident, I rate it as less likely. We can generalize this method to describe what it is to be over- and under-confident in beliefs more generally: overconfidence is a tendency or disposition to rate one's beliefs more likely to be true than the evidence merits; underconfidence is a tendency to do the opposite. In this sense, the philosopher who doesn't believe her work because of underconfidence has a calibration problem—she isn't apportioning her degrees of belief correctly.

If the (lack of) confidence at issue in believing our published claims is primarily a matter of mismatch between the claims authors commit to in their papers and the claims they believe their arguments support, authors can solve the problem of underconfidence by calibrating their claims more carefully to reflect what they think their evidence does support, or to more accurately report their degrees of belief.

But confidence can also be self-referential: it can pick out an attitude the agent takes, not towards her beliefs, but towards herself. This is the sense that contrasts with ‘self-doubt’ or ‘insecurity,’ and it’s not captured by our epistemic conception of confidence. If this is the root of an author’s doubts, the solution is not so simple, because lack of this type of confidence concerns the author’s ability to make the very judgments involved in calibration—an author who lacks this kind of confidence can’t simply rewrite her work to report the degree of belief she has in her arguments, because she lacks precisely the kind of confidence that makes her think she’s a good judge of such calibrations.

The kind of confidence I’m describing is something like what Jones (2012) describes as “intellectual self-trust.” My goal is not to outline a detailed conception of it here, but to show how it diverges from conceptions of confidence in terms of epistemic reliability or credence. Jones is adamant that self-trust is not a “purely cognitive” phenomenon; it is not simply a matter of beliefs about our reliability, or the level of confidence we have about a belief or beliefs in a given domain. Instead, self-trust is “a stance that an agent takes towards her own cognitive methods and mechanisms, comprising both cognitive and affective elements.” (2012: 238)

This stance affects our perceptions of error as well as our behaviors. Those who lack self-trust will be prone to doubt; they’ll focus on the possibility of error; they’ll be hesitant to rely on their own judgments. Perhaps most importantly for our purposes, they’ll be hesitant to “assume the asserter’s burden”:

In the act of asserting you present yourself as having been responsible in the use of your epistemic capacities and as willing to stand by the truth of what you assert... You present yourself as an equal co-participant in a shared practice of inquiry... To assert is simultaneously to claim credibility. The self-trusting are willing to claim credibility in a way that the self-distrusting are not. (Jones 2012: 244)

Authors’ willingness or hesitance to claim credibility is at the heart of the issue here. While it’s not wrong to frame the debate over PWB in terms of belief, we’d do well to distinguish cases where the author fails to believe her arguments because she is hesitant about the *evidence* from



cases where she fails to believe her arguments because she is hesitant about *herself*. As we've seen, the former can be addressed by hedging and qualifying a paper's claims, but the latter is harder to remedy. And because it's hard to distinguish these cases in practice, we should refrain from endorsing a belief norm about publishing and allow PWB.

Rejecting a belief norm may benefit insecure or self-distrusting authors, and offer a corrective to the intellectual combativeness in philosophy. It does so not by lowering the bar for publishing on individual occasions or making it easier for authors to game the system, but by lowering barriers for entry into the community of publishing. In other words, in the best kinds of scenario, PWB allows authors to enter into the publishing community in a way that helps them build the self-trust, and the confidence, that leads to publishing *with* belief. To be self-trusting is to stand by one's methods, conclusions, and arguments when others dispute them, and the act of doing this in turn increases our confidence in our epistemic abilities. The remedy for deficient self-trust requires us, "to come to have the right affective attitude towards our cognitive competence in a domain... it might be done by ingraining recognition of the competence we had all along into new habits of feeling, sensitivity to reasons, responses to disagreement and so forth."

Importantly, this is difficult to do as an individual; it is likely to require others to bolster our confidence in our abilities, to point out the sources of our self-doubts, and so on. As Jones observes, "If the problem of excessive self-distrust is social, then so must be its solution." (2012: 249)

I've already discussed my suspicion that our intuitions are muddled by examples that involve epistemically sketchy agents. So let me conclude by presenting another example:

Rayna is extremely conscientious, but insecure. She worries about speaking in seminars, always expecting there's some flaw in her claim that she hasn't spotted yet but which will be obvious to her classmates. When her advisor suggests she publish one of her dissertation chapters, she hesitates, unsure whether it's good enough—her advisor is probably just being nice. Nonetheless, she publishes the chapter. When it's out, it prompts a series of critical replies, which further undermine Rayna's confidence in the original argument. But because her newly-secured job depends on it, she continues to

defend the argument, even though she doesn't have much confidence in it. (If it matters, the original argument is correct).

One response is that Rayna sounds both pathologically insecure and poorly suited to academia. Perhaps. She might be an outlier; she might be a more common character than we think. But the case presents us with the question of who the norms of publishing are for, and against which kind of background we should evaluate them: philosophy as it is, or philosophy as it should be?

## **6. Two Pictures of Philosophy**

What's at issue here isn't just a view of publishing, but a view of philosophy itself. The two sides to the debate are operating with two different models of philosophy, in mind, and two different approaches to identifying its norms. The arguments for PWB I've offered work best against the backdrop of a 'non-ideal' model of philosophy. To wit: I've noted that insecurity and self-doubt motivate philosophers as much as truth and justification; my point about the reasons for not qualifying one's arguments suggests that which articles get published isn't just a matter of merits. I've pointed to the outsize importance of publication for getting and keeping a job, and suggested that if we object to PWB the solution is to reform our philosophical practices, not adopt a norm against it: "The process of publishing philosophy is subject to so many contingencies we ought to eliminate them wherever we can, especially where these affect the prospects of early-career philosophers whose beliefs are not yet calcified." (645)

It might sound odd to say that the critics of PWB are operating under an 'ideal' model of philosophy, since so many of their cases involve authors committing epistemic sins of varying severity. But the idealization is evidenced partly by the very assumption that (lack of) belief is apportioned to (lack of) evidence. So, the criticisms of PWB assume that an author's lack of confidence in his belief reflects a lack of confidence in his evidence. But the discussion in the previous section reveals that this is not the case.

Nor is the process by which we develop confidence in our ideas and arguments as individualistic as some of the examples make it seem. Kidd (2016: 400) observes, "It is an ugly truth that many able philosophers find that their confidence is damaged and sometimes even destroyed by their

experiences of argumentation.” Our confidence in any particular idea or argument is the product not just of the evidence for or against that idea, but of our experience with argumentation in seminar rooms, conferences, etc. And to the extent that an agent has “a history of experiences of unjust confidence depression,” (ibid.) her confidence may be less well-calibrated to the status of her evidence.

If we’re thinking about ideal philosophy, we may not notice that the way we engage in it, in practice, affects individuals differently; we may not notice that individuals bring their own sets of experiences and insecurities to it. We may also not attend to the way non-epistemic factors intersect with judgments about the merits of one’s arguments and the suitability of one’s papers for publications. To be fair, all the critics of PWB acknowledge and discuss these issues. But the norms they advocate are designed for a philosophy that has addressed them. For example, on the question of whether hedging would affect chances at publication, Buckwalter writes, “Belief encourages and motivates researchers to correct the research record.” That may be true in ideal, epistemically virtuous cases, but we can also identify prominent cases of researchers reluctant to acknowledge shortcomings in their published work for reasons that seem also to stem from belief.<sup>6</sup>

Fleisher’s ERC/ARC distinction also assumes an idealized picture of philosophy, because it ties our implementation/application of the norm to an idealization: a distinction that’s difficult to detect (if not nonexistent) in actual philosophical practice. This is the kind of idealization Mills (2005: 166-7) calls “ideal-as-descriptive-model”: it’s a picture of how philosophical papers work that makes “simplifying assumptions,” because the model being used deviates from the actual thing—or in this case, practice—being modeled. Mills points out that sometimes, the model itself represents our ideal of how things ought to work. It’s not clear whether Fleisher means the ARC/ERC distinction to be ‘merely’ descriptive or to be an endorsement of how philosophical papers ought to work, but either way the distinction relies on abstraction from the complications of actual philosophical argumentation. Above, I argued that a claim can play both roles (ERC

---

<sup>6</sup> For example, see Zimbardo’s response to questions about the Stanford Prison Experiment here: <https://www.vox.com/science-and-health/2018/6/28/17509470/stanford-prison-experiment-zimbardo-interview>; see also <https://www.nationalgeographic.com/science/article/failed-replication-bargh-psychology-study-doyen>. For a discussion of willingness to admit wrongness in published work, see Fetterman et al 2019.

and ARC) simultaneously. In theory, that's not a dealbreaker, but it raises the question of which norm ought to govern ambiguous claims, or papers containing both types of claims—a strong BNP-type norm, or a highly permissive PWB-type norm? My answer is the latter. We ought to prioritize helping the less confident over our worries about Malicious Deceivers (Plakias 2019: 638, partly because of considerations about non-ideal influences on philosophical belief. Critics of PWB disagree, but I think this is because they are looking at a different issue: the question of which norm would govern an ideal philosophy and its publishing practice. Here I have in mind the kind of idealization Mills calls 'ideal-as-idealized-model': the philosophical practice we would like to have, enacted by the philosophers we would like to be. We would like a philosophical practice that is influenced by epistemic merit, attention to arguments, and shared goals of inquiry directed at truth, not prestige-- even if this comes at the expense of individually publishing conclusive claims or exciting results. We would like to be agents whose degrees of belief are calibrated to evidence, not influenced by cognitive biases. But this is not the practice we have, and our norms should reflect the realities. In adopting norms for a less-than-ideal philosophy, we help the less confident among us develop their own intellectual self-trust, which may in turn broaden the contributions to our shared philosophical discourse.

I say 'may,' intentionally. Does PWB also leave us vulnerable to being taken advantage of? Yes. But the malicious or epistemically lazy among us are unlikely to be deterred by adopting a stronger norm of BNP. I've pointed out various ways we might judge and sanction the practices involved in cases like these, ways that don't require adopting a norm like BNP and are consistent with PWB. These forms of criticism—involving more general epistemic obligations and virtues like diligence, not misrepresenting the extent of one's research, using reliable methods, etc.—also restrict what authors are able to publish, since they require authors to have good evidence and justification for the claims they make in print. This is the flip side of Buckwalter's argument that belief curates the research record: the requirement for evidence itself curates the record, independently of belief. Belief will usually come along for the ride, but not always, and when it fails to, that may be the result of a deficit in self-trust which publication can help remedy.

## **7. Conclusion**

The debate over the norms of publishing is developing rapidly, and journals and reviewers become increasingly self-reflective about their procedures.<sup>7</sup> I haven't surveyed all the issues here, choosing instead to focus only on publishing within philosophy (as opposed to other academic fields) and offering a gloss on what's at issue. I've offered a diagnosis of what's at issue in the debate over publishing without belief that illuminates the connections between publishing, confidence, and self-trust. While philosophers have (perhaps unsurprisingly) focused on confidence as an intellectual attitude, we've been less quick to acknowledge the emotional underpinnings of self-confidence, and the emotional aspects of writing, publication, and philosophical argumentation more generally. I've suggested that the norms of publishing should depend on philosophy as we find it and as we would like it to be; we should prioritize norms that will make philosophy accessible to those whose confidence may have been undermined by its tendency towards aggressive argumentative practices. In doing so, we lower the barrier to entry to the profession while helping foster confidence. Allowing PWB may make the profession vulnerable to the epistemically malicious, but it also makes us more inclusive of those vulnerable to self-distrust.

### Works Cited

Bright, L.K. (2017). 'On Fraud.' *Philosophical Studies* 174 (2):291-310.

Buckwalter, W. (forthcoming). 'The Belief Norm of Academic Publishing.' *Ergo*.

Dang, H. & Bright, L.K. (2021). 'Scientific Conclusions Need Not Be Accurate, Justified, or Believed by their Authors.' *Synthese* (3-4): 1-17.

---

<sup>7</sup> One example: some journals have begun including an instruction asking that when composing reviews, reviewers put themselves in the author's position and ask themselves how they would feel if receiving the review.

Fetterman, A., et al. (2019). 'On the willingness to admit wrongness: Validation of a new measure and an exploration of its correlates.' *Personality and Individual Differences*. 138(1): 193-202.

Fleisher, W. (2020). 'Publishing Without (Some) Belief.' *Thought: A Journal of Philosophy* 9(4): 237-246

Jones, K. (2012). 'The Politics of Intellectual Self-Trust.' *Social Epistemology* 26(2): 237-251.

Habgood-Coote, J. (forthcoming). 'What's the Point of Authors?' *British Journal for the Philosophy of Science*.

Kelly, T. (2005). 'The Epistemic Significance of Disagreement.' In John Hawthorne & Tamar Gendler (eds.), *Oxford Studies in Epistemology, Volume 1*. Oxford University Press. pp. 167-196.

Kidd, I. (2016). 'Intellectual Humility, Confidence, and Argumentation.' *Topoi* 35: 395-402.

Lackey, J. (1999). 'Testimonial Knowledge and Transmission.' *The Philosophical Quarterly* 49(197), 471-490.

Levy, N. (2022). 'In Trust We Trust.' *Social Epistemology* 36(3): 283–298.

Levy, N. (forthcoming). *Philosophy, Bullshit, and Peer Review*. Cambridge University Press.

Melnikoff, D. & Strohminger, N. (2020). 'The Automatic Influence of Advocacy on Lawyers and Novices.' *Nature Human Behavior*, 4: 1258-1264.

Mills, C. (2005). "'Ideal Theory" as Ideology.' *Hypatia* 20(3): 165-184.

Plakias, A. 2019. 'Publishing Without Belief.' *Analysis* 79 (4):638-646.

Plakias, A. 2020. 'Some Probably-Not-Very-Good Thoughts on Underconfidence.' *Ethical Theory and Moral Practice* 23 (5):861-869.

Sarihan, I. 2022. 'Problems with Publishing Philosophical Claims We Don't Believe.' *Episteme*:1-10.