

Essays in Philosophy

ISSN 1526-0569 | Essays in Philosophy is published by the Philosophy Documentation Center, in cooperation with Pacific University

Volume 21, Issue 1-2 (2020)

The Philosopher as Moral Activist: A Call for Ethical Caution in Publication

Kyle York
Lingnan University, Tuen Mun, Hong Kong

Abstract

It is normal to think that philosophers' first dedication is to the truth. Publishers and writers consider ideas and papers according to criteria such as originality, eloquence, interestingness, soundness, and plausibility. I suggest that moral consequence should play a greater role in our choices to publish when serious harm is at stake. One's credence in a particular idea should be weighed against the potential consequences of the publication of one's ideas both if one turns out to be right and if one turns out to be wrong. This activist approach to philosophical writing combines moral concern with epistemic humility.

Essays Philos (2020)21: 46-75 | DOI: 10.5840/eip2020211/24
Correspondence: kylenyork@gmail.com



© 2020 York. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

I. Introduction

Many philosophers (especially ethicists) have wanted to bring about some change in the world with their writings.² Philosophers also occasionally choose not to publish some work in fear that it would cast their other works in a negative light, such as Jeremy Bentham when he hid away his writings calling to end the persecution of homosexuality. However, to systematically approach these sorts of issues is far less common. I can think of many philosophers, however, who should have considered this question.

The most prominent criteria for publishing in philosophy are things such as interestingness, originality, plausibility, soundness, eloquence, scope (such as broadness of interest), and the potential significance or groundbreaking quality of conclusion. Some of these criteria are hopefully reliable indicators of truth. Others are indicators of the (usually theoretical) importance of that truth. For the most part, the moral consequences of publishing a particular work are not adequately considered as a criterion for publication. Of course, some work could conceivably excel in most respects yet argue for a conclusion that is ‘off the table’, such as the permissibility of slavery or the inferiority of some race or gender. There are also discussions about whether or not we should platform potentially harmful ideas, whether we should engage with them, and whether or not we should pursue potentially harmful lines of inquiry.³ These are exceptions, however, rather than the rule.

To remedy this absence of moral consideration, I am going to propose two principles to use for publishing in philosophy. Firstly, if an argument (or body of work) would excessively risk causing (foreseeable) serious harm to others, then one has a moral duty not to publish that argument unless doing so carries comparable moral benefits. Call this the ‘negative principle’. By ‘serious harm’ I mean such things as death, injury, emotional

¹ I would like to give special thanks to Derek Baker, Andrea Sauchelli, Merily Salura, and the anonymous reviewers of this essay for their useful notes and feedback. I also give thanks to some various (and mostly anonymous) philosophers online for their helpful thoughts, Ramona Ilea for her patience and helpfulness as the editor, and the Hong Kong Research Grants Council for their support.

² Kasper Lippert-Rasmussen, “The Nature of Applied Philosophy,” in *A Companion to Applied Philosophy*, ed. Kasper Lippert-Rasmussen, Kimberley Brownlee and David Coady (Hoboken, NJ: John Wiley and Sons, Inc., 2016). Cited in Wendy Rogers, “Bioethics and Activism: A Natural Fit?” *Bioethics* 33, no. 8 (2019): p. 883.

³ For example, see Justin Weinberg, “Recognizing Gender Critical Feminism as Anti-Trans Activism (Guest Post),” *Daily Nous*, August 6, 2019, <http://dailynous.com/2019/08/06/recognizing-gender-critical-feminism-anti-trans-activism-guest-post/>.

trauma, strong oppression, and the deprivation of basic rights and needs.⁴ The second principle is this: if an argument (or body of work) would sufficiently raise the likelihood of preventing (foreseeable) serious harm to others, then so long as one has some reasonable credence in one's arguments (in that one is not outright lying or falsifying information), one has a moral duty to publish that argument unless doing so carries comparable moral drawbacks.⁵ Call this the 'positive principle'. For the sake of simplicity, I will mostly focus on arguments, but my points will also apply to bodies of work. I leave it open as to whether the harms in question must be in violation of another's rights.⁶

I have chosen to focus on strong negative harms because such impacts are morally more significant than what Joel Feinberg refers to as "offenses, hurts, and... the various unhappy and unwanted physical and mental states that are not states of harm in themselves."⁷ Additionally, they are morally more significant than harms that constitute setbacks to our wants, needs that are not basic, and 'ulterior interests' (interests that are related to our personal projects and goals). The sorts of harms to welfare that I am interested in are more serious than setbacks to ulterior interests in part because "when they are blocked or damaged," one's "more ultimate aspirations are defeated too."⁸ Plus, if the negative principle were to include things like hurts and setbacks to ulterior interests, the scope of moral permissibility for publication will become too narrow. For example, some publications may set back the frivolous but deeply held interests of some privileged group or greatly offend some intolerant person's sensibilities, but we would not therefore want to say that the writer morally ought not have published it, even if the publication had no moral benefits. Keep in mind, however, that I am only interested in discussing what we

⁴ By 'basic needs' I mean something like what is meant by 'course-of-life' needs in David Braybrooke, *Meeting Needs* (Princeton, NJ: Princeton University Press, 1987), cited in Gillian Brock and David Miller, "Needs in Moral and Political Philosophy," *Stanford Encyclopedia of Philosophy* (Stanford University, April 11, 2019), <https://plato.stanford.edu/entries/needs/#NeedMoraObli>.

⁵ It might be asked whom exactly I have in mind when I say, "one has a moral duty." To whom are my principles meant to apply? My points certainly apply to writers and may also apply to editors, unless there is a reason that editors should be exempt.

⁶ In Shelly Kagan, *Normative Ethics* (Boulder, CO: Westview Press, 1998), p. 170, Kagan suggests that in the broadest sense of 'rights', "to say of something that it has rights is only to say that it has moral standing—to say that it *counts* from the moral point of view." Kagan sees this in part as a simple shift in emphasis; if we ought not harm something, then in this sense that thing has a right against being harmed (*ibid.*, p. 171).

⁷ Joel Feinberg, *The Moral Limits of the Criminal Law Volume One: Harm to Others* (New York: Oxford University Press, 1987), p. 47.

⁸ *Ibid.*, p. 37.

morally ought to do and not what we should have a legal or social right to do. We should have the legal and social rights to do all sorts of things that we nevertheless morally ought not do.

I should also point out that I only mean to supplement our current criteria in publishing philosophy with the positive and negative principles as opposed to replacing them completely. What this entails, however, will depend on larger questions that I will not have space to address. Firstly, it will depend on more general ethical considerations that will tell us how to interpret and implement the principles. For example, when the positive principle weighs towards publishing some work, one might interpret this as obligatory or supererogatory, depending on one's more general views about topics such as the ethics of preventing harm. Secondly, it will depend on considerations about the place of moral values and obligations when weighed against other sorts of values and obligations.⁹ For example, even if the negative principle morally obliged one not to publish, some may agree with Bernard Williams that moral obligations cannot always trump other sorts of considerations.¹⁰ For my part, I will simply be considering the question of publishing from a moral point of view and limiting my considerations to what would be moral or immoral to publish.

I will first provide a basic argument against causing serious harm through publishing. Next, I will show that there are problems with prohibitions against causing harm in publishing, and why a principle against risking harm is preferable. I will suggest a method for assessing and making decisions in the face of moral risk. Finally, I will turn to the positive principle before considering some important objections. Due to limited space, I will mostly rely on the commonsense positions on our duties and rights but will develop on these ideas when necessary.

I will conclude that my principles will cast (at least practical) philosophers in the role of a moral activist. How much of an activist my principles will recommend philosophers become will depend both on one's favored definition of 'activism' and how demanding one thinks the positive principle should be.¹¹ On the weak end of the spectrum, phi-

⁹ See, e.g. Susan Wolf, "Moral Saints," *The Journal of Philosophy* 79, no. 8 (1982): p. 419.

¹⁰ Bernard Williams, "Morality, the Peculiar Institution," in *Ethics and the Limits of Philosophy* (London: Routledge, 2011).

¹¹ What exactly qualifies as activism is a contested issue in the literature, so it is possible that what I am advocating will not count as activism under all definitions. I am not here to resolve this issue of what qualifies as activism. At any rate, it is enough for my purposes that philosophers are at least

losophers would be activists in the sense of prudently leaving overly-risky arguments unpublished or aiming to shift each other's credence towards attitudes that seem to have important and positive ethical consequences.¹² In the middle of the spectrum, philosophers would also aim their arguments towards the general public and policy makers in the most fitting and effective mediums they could find.¹³ On the extreme end of the spectrum, philosophers would totally focus their philosophical work towards minimizing serious harm to others.¹⁴ In all of these cases, philosophers would let moral considerations significantly influence their publishing practices, not merely contemplating truths but trying to change the world.

II. Causing Harm

Let us now turn to a basic defense of the negative and positive principles. Imagine the following case:

Case 1: Petra has just written an argument that it is morally permissible to eat meat. She publishes the argument, and Susan finds her arguments so convincing that she abandons her vegetarian diet. Susan consumes 52 animals over the next year before she is converted to vegetarianism again. As it turns out, eating meat really is in violation of animals' rights.¹⁵

In this case, Petra's act of speech has caused serious harm to other beings against their rights. If we are prone to judge that Petra has nonetheless acted permissibly, why? It could be because Petra's argument seemed true to her, and she intended no wrong. In the next section, it will become clear that Petra could still be accountable for the risk she

called *towards* the activist end of engagement on the spectrum between 'pure' theory and activism, and that my principles call for activism under some established conceptions of the term. For more on the spectrum of engagement, see Heather Draper, "Activism, Bioethics and Academic Research," *Bioethics* 33, no. 8 (2019).

¹² This is sometimes called 'standard activism.' See Michelle Dempsey and Matthew Lister, "Applied Political and Legal Philosophy," *A Companion to Applied Philosophy*, July 2016, p. 319. Cited in Draper, *op. cit.*, p. 862.

¹³ Dempsey and Lister (*op. cit.*, p. 315) think of this as 'extreme activism,' though Draper (*op. cit.*, p. 862) sees their definitions as too "passive and all-inclusive."

¹⁴ This is closer to what Draper (*op. cit.*, p. 863) thinks of as 'extreme activism.'

¹⁵ This figure is based on those found in Noam Mohr, "Average and Total Numbers of Animals Who Died to Feed Americans in 2008," (United Poultry Concerns, August 2009), <https://www.upc-online.org/slaughter/2008americans.html>. I excluded the 218 shellfish that the average American eats in a year.

took, but even setting that point aside, Petra's intentions do not matter here. For even if Petra is not morally blameworthy for her actions, her ignorance would not make her actions morally permissible. All sorts of immoral acts have been done by people under the impression that they were doing the right thing.¹⁶

It might be that Petra's article only caused harm indirectly.¹⁷ After all, Petra just published her thoughts. It was Susan who harmed the rights of the animals by eating them for pleasure. This seems to rely on the premise that any harm is indirect when it is intermediated by a third party (between the offending and harmed agent). But if this is so, it is unclear how any act of speech could cause direct harm. However, we still would want to say that white supremacists ought not incite violence against oppressed social groups. The fact that intermediary parties are the ones carrying out the physical harm is inconsequential in this case, so we would need some special reason to think that Petra's case would be different.¹⁸ A more proper assessment of indirect harm is in terms of diffused social effects and direct harm in terms of immediately inspiring action. But I also do not see why this distinction should make a difference in Petra's case. Petra may have contributed to a malformed moral environment, but more to the point, she has convinced Susan to eat meat. Moreover, it is possible that Petra *does* commit a direct or unmediated harm towards animals by undermining the moral standing to which they are (*ex hypothesi*) entitled.¹⁹

There are a few more reasons why we might consider her actions permissible, related to the fact that she was engaged in an ongoing public debate about morality and animal rights. In the objections section, I will address this sort of issue in more detail. For now, I will simply say that this fact alone does not excuse Petra's actions. Whether or not he is blameworthy for doing so, Kant ought not have justified slavery, even if his justification was part of an ongoing intellectual tradition or debate about applied ethics. Thus, it

¹⁶ On issues such as this, as we will later see, judgments from subjective and objective norms will differ.

¹⁷ See, for example, Edward Royzman and Jonathan Baron, "The Preference for Indirect Harm," *Social Justice Research* 15, no. 2 (2002); and David van Mill, "Freedom of Speech," *Stanford Encyclopedia of Philosophy* (Stanford University, May 1, 2017), <https://plato.stanford.edu/entries/freedom-speech/#ResHarPri>.

¹⁸ This same point applies to J. S. Mill's limitation of speech only insofar as it directly harms the rights of others. If this were our picture of direct harm, it would be hard to make sense of why Mill would grant that accusations that corn dealers are starving the poor "may justly incur punishment when delivered orally to an excited mob." J. S. Mill, *On Liberty* (London: Longmans, Green, Reader and Dyer, 1869), https://en.wikisource.org/wiki/On_Liberty.

¹⁹ See: Jeremy Waldron, *The Harm in Hate Speech* (Cambridge, MA: Harvard University Press, 2014), p. 33.

seems that one may only judge Petra's publication against animal rights as morally permissible because one is oneself unsure about whether animals have rights. For although Petra is mistaken according to the omniscient narrator of the imagined example, in the real world, many doubt that she is mistaken.²⁰ Thus, we should now turn to the issue of risking harm.

III. Risking Harm and Normative Uncertainty

There are good reasons to doubt that the actual effects of a publication are necessary or sufficient to determine its moral status. Consider these three cases:

Case 2: Susan misunderstands Petra's argument against animal rights, which happened to be true, and becomes convinced that it is also permissible to kill and eat people. Susan then kills and eats people.

Case 3: Petra (as it happens, rightly) argues in favor of animal rights, but Susan comes to hold the opposite opinion because Susan irrationally hates the font and typeset that Petra's argument uses.

Case 4: Petra and James both publish the same basic argument that it is permissible to eat meat. Susan reads James' argument first and is already convinced (as it happens, wrongly) to abandon her vegetarian diet by the time she reads Petra's argument, which Susan finds just as convincing as James'.²¹

To establish what Petra ought to do, it seems better to consider the possible and likely consequences of her work.²² Cases 2 and 3 are possible, but both are so unlikely that it

²⁰ This makes sense in the context about the permissibility of making arguments, since the moral, social, and political arguments that we are most certain are wrong (such as those of Nazis) tend to be the ones we would be the most willing to consider immoral to make.

²¹ This example was inspired by an example about breaking a window in Christopher Read Hitchcock, "The Role of Contrast in Causal and Explanatory Claims," *Synthese* 107, no. 3 (1996). Cited in Jonathan Schaffer, "The Metaphysics of Causation," *Stanford Encyclopedia of Philosophy* (Stanford University, July 5, 2016), <https://plato.stanford.edu/entries/causation-metaphysics/>.

²² This is not to say that causing harm is not morally significant independently of risking harm. There is a lot of literature on causing and risking harm and I do not wish to take a position on the relative importance of these phenomena. But since, when it comes to publishing, cases of wrongfully harming all seem to also be cases of risking harm (and since not all cases of wrongfully risking serious harm are cases of causing it), focusing on risking harm instead of causing harm allows us to avoid condemning

would be strange to say that she ought not to have published. On the other hand, the risk at which Petra placed animals in Case 4 may be sufficient to establish that she ought not to have published her argument; the fact that Petra's argument went on to actually cause harm in Case 1 at most just makes it worse. Using likelihood and severity of consequences as a method of moral assessment is also more useful for making real-world decisions, where we are rarely certain about what the consequences of our writing will be.²³

As is apparent already, the simple fact that a publication carries risks is not enough to base moral judgments upon; a threshold for acceptable risk has to be established. In other words, some publication would be immoral if $P(h) \geq x$, where 'P' stands for probability, 'h' stands for serious harm, and 'x' stands for some threshold between 0 and 1.²⁴ Although it is not clear what x should be, this is not a special problem for my principles. Call this the 'objective version' of the negative principle. Without reasons otherwise, it is probably safe to say that x is the same as the threshold for the unacceptability of other sorts of risks. For the objective version of the negative principle, it does not matter whether the writer recognizes the risk, as long as the risk is reasonably foreseeable.²⁵

Our uncertainty, however, is not only about the empirical effects of our publications but the moral effects as well. In Case 1, Petra might have rightly predicted that her argument would likely cause someone to eat meat, yet neglected to properly consider that this could be a morally negative effect. After all, the moral status of animals is the very thing that Petra was questioning. Here, we might say that Petra was wrong for placing animals at risk whether or not we know that seriously harming animals is wrong. To do this, we would take Petra's normative uncertainty into account and show how she fails to act properly in terms of the subjective norms governing someone in her epistemic position. Call this the 'subjective version' of the negative principle. The severity of the risk involved here must be counterbalanced by Petra's credence in various moral evaluations

cases of accidentally causing harm and handle causally ambiguous cases such as philosophers who influence other philosophers whose arguments cause harm.

²³ Van Mill (op. cit.) thinks that because it is difficult to determine the actual effects of a publication, Mill's harm principle will make it difficult to limit speech in most cases. However, as David Brink notes, "In many cases all we could reasonably know is that a given action *risks* harm. Fortunately, this seems to be all that Mill requires." See David Brink, "Mill's Moral and Political Philosophy," Stanford Encyclopedia of Philosophy (Stanford University, August 21, 2018), <https://plato.stanford.edu/entries/mill-moral-political/#HarPri>.

²⁴ The value of x might change with the severity of the harm in question.

²⁵ The agent will be blameworthy only if she should have foreseen the risk, but the action will be impermissible regardless.

of the possible outcomes. When we are not sure what the right thing to do is, we can apply a framework of decision based upon the value of possible outcomes multiplied by the percentage of our credence in those outcomes.²⁶ Imagine this case:

Case 5: Petra has 90% credence that it is okay to eat animals for pleasure (normative uncertainty). She also believes that there is a 50% chance that Susan will be convinced and eat 52 animals over the next year (empirical uncertainty).

The empirical uncertainty is easy to handle. Fifty percent of 52 is 26, so the expected consequence of publishing will be the death of 26 animals. The normative uncertainty is harder. We can represent the possible moral outcomes like this:

	A: Animals matter	B: Animals do not matter
Publish	Very bad	Mildly good
Do not publish	Very good (or neutral)	Mildly bad (or neutral)

Recall that Petra has 90% credence that animals do not matter (and 10% credence that they do).²⁷ Even so, if the animals' lives are only ten times more morally important (in case A) than Susan's pleasure in eating them (in case A or B), Petra should not publish, since the expected value of publishing would be -0.1 and the value of not publishing would be either zero or 0.1:

	A: Animals matter (10%)	B: Animals do not matter (90%)
Publish (A+B= -0.1)	$-10 * 10\% = -1$	$1 * 90\% = 0.9$
Do not publish (A+B= 0.1 or 0)	$10 * 10\% = 1$ (or zero)	$-1 * 90 = -0.9$ (or zero)

If we replace the permissibility of eating animals with that of committing murder in this example, the plausibility of the principle becomes even clearer.

One may object that we should not select the option of highest expected value but may instead select the option that is recommended by the moral theory they have the most credence in. While William MacAskill, Toby Ord, and Ted Lockhart provide good rea-

²⁶ See William Macaskill and Toby Ord, "Why Maximize Expected Choice-Worthiness?," *Noûs*, 2018; and Ted Lockhart, *Moral Uncertainty and Its Consequences* (New York: Oxford University Press, 2000), p. 4.

²⁷ For the sake of simplicity, let us assume that Petra's credence in some belief that p is 100% minus her credence that she is wrong in believing that p.

sons why moral decision making should operate according to value (or choice-worthiness) maximization, which is usual in decision theory, for my purposes, it is enough to say that when an option involves sufficient risk (the sort that would be considered unacceptable in ordinary circumstances), we ought not select that option, even if our credence favors it. Consider the analogy of acting on empirical uncertainty. Imagine that I am serving food to guests at a dinner party, and I have 99% credence that my food is not poisoned. This amount of certainty is not sufficient to justify serving them the food because the stakes are so high. Of course, it would be unreasonable to demand that I ought to be entirely free of all doubt that my food has been poisoned before serving it (99% in such a case would be an alarmingly low credence). But the level of justified certainty writers and philosophers have in their own claims is rarely if ever going to be so high as to raise these sorts of objections.²⁸ Thus, we have two ways of judging Petra's choice to publish in Case 1. Objectively, she ought not have put animals at risk. But even from the perspective of Petra's own moral beliefs and credence therein, assuming it to be reasonably moderated, we can make a case that she ought not to have published.

One possible problem with this view is that it would not apply to philosophers with 100% credence in their views. However, we almost never have justified certainty about these sorts of things, so even if Petra feels certain about this issue, she should not. This should be intuitively obvious; how can Petra know for sure? There are also reasons to think this. For one thing, if certain epistemologists are right, expert disagreement and controversy on animal rights should lower Petra's credence in her views.²⁹ Setting this point aside, having 100% credence that animals do not matter, from a Bayesian perspective, prevents Petra from even being able to update her belief in light of new evidence.³⁰ One might also object that some argument could cause immediate harm to some group but improve the lives of its members in unforeseeable ways decades later. This is a common but unsuccessful objection to consequentialist ethics. If I see a drowning child, I ought to save her even if I cannot know that she will not grow up to be a serial killer

²⁸ It might also be objected that driving also carries expected risk to others but it is the sort of activity we find acceptable. One might make the case that drivers and pedestrians all consent to this risk, but children do not. However, if the risks of getting into a car crash were as high as 1% every time one got onto the road, driving around children would not seem so acceptable. If my principles were applied to most acts of driving, the risk of serious harm is probably minimized by the unlikelihood of a crash enough to be outweighed by the morally relevant benefits.

²⁹ See Bryan Frances, *Disagreement* (Cambridge: Polity, 2014). The issue of whether to adjust credence in the face of peer or expert disagreement is controversial, but my argument does not rely on it.

³⁰ See Darrell P. Rowbottom, *Probability* (Cambridge, UK: Polity, 2015): p. 164-166. Thanks also to Darrell Rowbottom for discussing this and a few issues regarding decision theory with me.

decades later. When we have no reason to believe our action will have an effect one way any more than the opposite way, the unforeseeable risks of both taking some action and not taking some action negate each other.³¹ What we should work with, when doing moral reasoning, are the risks we can foresee and have reason to believe exist.

Likewise, we should not be concerned that our foreseeable risks will usually cancel each other out. There is no reason to suppose that the stakes are going to be the same on both sides of every moral argument. In fact, in regards to the objective version, it would be absurd to suppose that an argument on every side of an issue will carry the same sort of risk that we usually take to be unacceptable, and regarding the subjective version, it would be absurd to suppose that every possible moral decision we make will have the same expected value. On issues where publishing an argument for and against some point of view both carry what would in ordinary circumstances be an unacceptable risk of serious harm, but not publishing at all would also fail to prevent those harms, one might simply want to do whatever carries the slightest risk of harm, whether that is publishing the least risky argument or refraining from weighing in altogether.³²

IV. Preventing Harm

There is a weak and a strong version of the positive principle. The strong version conceives of it as a perfect duty, whereas the weak version regards it as an imperfect duty. By ‘imperfect’ I just mean that the positive principle leaves us with options regarding how and when we discharge our duty.³³ There is also a very weak version of the positive principle, which regards acting on the principle as a supererogatory good, but I will ignore this version. Most of us believe that we have strong duties in certain circumstances to aid and prevent

³¹ See Matthew Rendall, “Discounting, Climate Change, and the Ecological Fallacy,” *Ethics* 129, no. 3 (2019): pp. 441-463. Cited by Samuel Scheffler, “Ethics Discussion at PEA Soup: Matthew Rendall’s ‘Discounting, Climate Change and the Ecological Fallacy,’ with a Critical Précis by Samuel Scheffler,” ed. Chike Jeffers, PEA Soup (Prindle Institute for Ethics, April 14, 2019), <http://peasoup.us/2019/04/ethics-discussion-at-pea-soup-matthew-rendalls-discounting-climate-change-and-the-ecological-fallacy-with-a-critical-precis-by-samuel-scheffler/>.

³² This is, of course, just one possible interpretation of the positive and negative principles and not one that I am committed to.

³³ Relatedly, Greg Moorlock suggests that it may be morally *better* if junior academics wait until they have established careers before engaging in activism, so they can do it more effectively. See Greg Moorlock, “Do Junior Academic Bioethicists Have an Obligation to Be Activists?” *Bioethics* 33, no. 8 (2019): pp. 922-930.

harm to others.³⁴ This speaks in favor of the strong version. Many of us, however, would also judge that this duty is only perfect in immediate emergency cases where there is no unreasonable risk to us. Of course, it is quite possible that we have these same duties when those in need are more spatially or temporally remote.³⁵ However, for the weak principle, it is enough to rely on the commonsense intuition that it is morally good to prevent harm in non-emergency cases, even if we are not violating a moral obligation when we do not prevent it, and that we would be blameworthy for never preventing such harm (even remote harm) despite ample opportunity to do so. I suspect that some combination of the weak and strong positive principle is true, depending on circumstances such as how high the stakes are.³⁶ The stakes will be a factor both of likelihood of success and the scope and severity of the harm involved.³⁷ It is also worth noting that certain philosophers, *qua* philosophers, may have special obligations towards harm prevention. This could be because, for example, certain philosophers (such as ethicists and social philosophers) have a specially trained capacity to identify “situations of moral trouble”.³⁸ It could also be because sometimes philosophers’ voices have legitimacy in places where marginalized people’s do not.³⁹

Just as we ought to prevent harm, we ought to take actions that raise the likelihood of preventing harm. As with the negative principle, we can apply the positive principle in a subjective or objective way. Assessing arguments in terms of likelihood of harm prevention also carries theoretical advantages over assessing arguments in causal terms alone, just as we saw with the negative principle. We can assess what we ought to do by the objective likelihood of our success according to a moral theory that is taken for granted.

³⁴ See also: Feinberg, *op. cit.*, chapter 5, especially sections 7-8.

³⁵ See, e.g., Peter Unger, *Living High and Letting Die: Our Illusion of Innocence* (New York, NY: Oxford University Press, 2011).

³⁶ Another reason to think that the strong positive principle is not usually true is that if we had perfect duties to create publications to prevent serious harm whenever possible, we would probably also have perfect duties to prevent serious harm in more effective ways than writing philosophy articles whenever possible, which would override our duties to write helpful philosophy articles.

³⁷ If, as is commonly thought, our constraints not to harm others are limited by certain thresholds depending on the stakes involved, it is probably also the case that the ordinary permissiveness of positive duties is constrained by certain thresholds. For more on thresholds, see Kagan, *op. cit.*, pp. 78-84.

³⁸ Jackie Leach Scully, “The Responsibilities of the Engaged Bioethicist: Scholar, Advocate, Activist,” *Bioethics* 33, no. 8 (2019): p. 872.

³⁹ Scully, *ibid.*, p. 878. See also: Deborah Zion, “On Beginning with Justice: Bioethics, Advocacy and the Rights of Asylum Seekers,” *Bioethics* 33, no. 8 (2019): pp. 890-895.

Alternatively, we can assess what we ought to do according to alternative moral theories based on our credence in those theories and the stakes involved according to each theory. For example, imagine the following case:

Case 6: Petra has 10% credence that affluent people have a moral duty to donate 10% of their extra income to buy mosquito nets to protect impoverished people from malaria (a serious harm). Let us assume for simplicity that she knows her paper will convince Susan to donate about \$50, enough money to buy mosquito nets for 50 people.⁴⁰ Otherwise, Susan will go out for a nice dinner.

If we assume (conservatively) that protecting 50 people from malaria is only 10 times more morally valuable (or even valuable *simpliciter*) than Susan's dinner, we can determine the expected value of Petra's publication:

	A: Susan has a moral duty to the poor (10%)	B: Susan has no moral duty to the poor (90%)
Publish (A+B= 0.1)	$10 * 10\% = 1$	$-1 * 90\% = -.9$
Do not publish (A+B= -0.1 or 0)	$-10 * 10\% = -1$ (or 0)	$1 * 90 = .9$ (or 0)

This is an extremely simplified version of how this reasoning would actually work, but the basic point should be clear: even if Petra has very low credence in a certain idea, she still ought to publish it if it has a positive expected value regarding an issue where the prevention of serious harm is at stake.⁴¹ Since credence as low as 0.1 that p probably counts as disbelief that p, this would imply that philosophers should publish assertions they do not believe. Some might take issue with this and wish to limit the application of my principle to cases where the writer has at least 0.5 credence in her assertions. This may be a sensible limit, but without convincing reasons to set it, I do not see why our

⁴⁰ "Why Nets?" The Against Malaria Foundation, accessed October 20, 2019, <https://www.againstmalaria.com/WhyNets.aspx>.

⁴¹ An anonymous reviewer has pointed out that epistemologists who defend the knowledge norm of assertion (that one should only assert p flat out if one knows that p) or weaker but similar variations of this norm may take issue with this particular application of my principles. While I do not have space to address the literature on this topic, this conflict might simply be a case of moral and epistemic norms pulling us in two different directions. Those who wish to avoid this sort of conflict may wish to add a proviso that when a philosopher is arguing for some proposition that she has low credence in, she should hedge and point out that she does not have full credence in the proposition. For a discussion of epistemic standards for philosophers in the public sphere, see Tom Shakespeare, "When the Political Becomes Personal: Reflecting on Disability Bioethics," *Bioethics* 33, no. 8 (2019): pp. 918-919.

notions of belief and disbelief should play a special role in determining acceptable levels of credence. Thus, for my purposes, I will leave my principles as they stand.

The positive principle, perhaps surprisingly, can also be the result of our negative duties to refrain from doing harm. For example, our obligations to help alleviate global poverty, Thomas Pogge argues, come not just from a duty to prevent harm but to refrain from harm since the corruption and misrule that contribute to poverty in developing countries are encouraged and imposed on those countries by the global institutional order to the benefit of the economically and politically powerful in affluent countries.⁴² Publishing works that are likely to prevent harm to those in developing countries in this case would be the discharge of a negative duty. This same kind of point could apply to advocating for animals and various socially and economically marginalized groups.

V. Convincing vs. Shifting Credence and Other Kinds of Serious Harm

It is not necessary to be able to convince some audience about some positive assertion for my principles to apply; it is sufficient to be able to shift their credence enough when a risk of serious harm is at stake.⁴³ But what does it mean to shift someone's credence 'enough'? I have yet to address the question of works that are not the sole cause but a contributing cause of some serious harm. For simplicity, my focus has been on cases of risk where the risking factor would (as in Case 1) or could (as in Case 4) be the only relevant causal factor when serious harm is done or prevented. However, not every argument that increases the risk of some serious harm will be a sufficient cause of the reader's change of attitude. Arguments that are not sufficient causes could still increase the risk of serious harm in a way that we would otherwise consider unacceptable. 'Enough', then, refers to this threshold of acceptability. For example, imagine this case:

Case 7: Petra publishes an argument that it is acceptable to eat meat. The paper has a high risk of decreasing Susan's credence in her (correct) vegetarian convictions by 10%. However, Susan's credence would have to change by over 75% for her to abandon her (morally right) vegetarian practices.

Petra has still probably done something she ought not have done here. For it might be that Susan will go on to read some other paper that decreases her credence in her vegetarian convictions by 66%, but would not have gone on to eat meat if it were not for

⁴² See Thomas Pogge, *World Poverty and Human Rights* (Cambridge: Polity, 2008).

⁴³ While I have for the sake of simplicity focused on arguments here, it is more likely that substantial bodies of work will have an effect on the overall credence of a community.

Petra's paper. As an analogy, imagine this situation:

Case 8: Assume that buying meat is morally wrong if and only if it has harmful consequences for animals (e.g. if it causes the grocery store to put in a new order from the meat suppliers). If I buy a steak, this decreases the store's stock by 10%. If no one else buys steak by the end of the day, all of that day's steak is thrown away. However, if the store's stock of steak is decreased by 75% before they close that day, they will put in an order for more steak.⁴⁴

In this situation, I ought not to buy a steak unless I am quite sure that no one else will buy any steak. If I buy one, I have contributed to a threshold that, if reached, would cause serious harm. And if this is true, then it is also probably true that Petra ought not to have published in Case 7.

This should broaden our view of what constitutes risk. The point also applies to the positive principle, which would endorse making an argument that contributes to shifting the credence of readers towards some threshold which, if reached, will prevent a serious harm. With these points in mind, one should see that the scope of philosophical arguments to which my principles apply is quite broad; they need not be convincing but merely credence-shifting enough.⁴⁵

We have now reached the end of the main body of my argument. Before we move on to objections, I should address a final concern: what about a work that makes likely the prevention of serious foreseeable harm while at the same time risking serious foreseeable harm? I leave the interactions of my principles for others to decide. Consequentialists may prefer whichever option avoids the greater total amount of harm, while deontologists may prefer restraints on risking harm that override preventing it.

⁴⁴ This case was inspired by one from Teemu Toppinen, "The Ethics of Dietary Choice: Does it Matter?" *The Philosophy of Food* (class lecture, University of Helsinki, Helsinki, Finland, May 4, 2018).

⁴⁵ Moreover, some arguments do not necessarily harm by changing anyone's credence in some viewpoint but by putting something 'on the table,' making it the subject of public discussion, and (sometimes) imposing it into the spaces of vulnerable people. Arguments about the validity of trans people or permissibility of rape, for example, may risk serious harm in just this way. See also Elizabeth Barnes, "Arguments That Harm - and Why We Need Them," *The Chronicle of Higher Education* (The Chronicle of Higher Education, February 18, 2018), <https://www.chronicle.com/article/Arguments-That-Harm-and/242543>.

VI. Objections⁴⁶

Philosophers do not make a difference in the first place.

If philosophy hardly makes practical differences in the first place, then the positive and negative principles would have almost nothing to which they could apply. After all, even among those contemporary philosophers who achieve fame in their field, most remain largely unknown to the public. As for the rest of us, as Michael Huemer has put it,

Quite a bit of intellectual talent and energy is being channeled into producing thousands upon thousands of papers and books that hardly anyone will ever read or want to read. These articles and books are written almost entirely for other academics working in the same sub-sub-sub-specialization that the author works in...mainly...so that the readers can cite those articles in their own articles.⁴⁷

There are a couple of ways to respond to this concern. The first way is to accept pessimism about the vast majority of philosophers having any significant practical effects, at least where publishing in journals is concerned. Even in this case, we should also keep in mind that the positive and negative principles may not only apply to publishing philosophical ideas and arguments in journals but in mass media, social media, educational material for professional fields (such as medical ethics), policy avocation, think tanks, and for cause-related research such as Effective Altruism.⁴⁸ This may even inject more lifeblood into the practice. After all, when we publish just for professional reasons, it may be true that, “most interesting ideas have already been discussed at great length. To continue to say ‘new’ things, we must divide our fields into increasingly specialized, tiny subfields.”⁴⁹ My principles may compel practical philosophers to do more of the kind of work that first made many of them love practical philosophy

⁴⁶ The structuring of this section was inspired by Philip Kitcher, “An Argument About Free Inquiry,” *Noûs* 31, no. 3 (1997): p. 298.

⁴⁷ Michael Huemer, “Publishing in Philosophy,” August 2015, <http://www.ow1232.net/?fbclid=IwAR3RYwaTjd1KRufz4Rw12DgBjayrIPziRSmbbl36PHTgHshnkM2cedulsMw>.

⁴⁸ See also: Draper, op. cit. and the Impact Ethics form, which aims to use “the tools of ethics to shock, press, crack, and chip society into a better place.” Chris Kaposy, Angel Petropanagos, and Françoise Baylis, eds., “About,” Impact Ethics, February 15, 2019, <https://impactethics.ca/about/>. Cited Draper, op. cit., p. 861.

⁴⁹ Huemer, op. cit.

with less of the professional pressure towards always having completely new ideas.⁵⁰

Alternatively, I could answer this worry by noting that even if we are not lucky enough to be read by one of the hundreds of thousands of students who enter philosophy programs every year (in the U.S. alone), a few other philosophers *do* still read us.⁵¹ If a philosophical argument or body of work is influential enough to sufficiently risk serious harm to just one person (or animal), my arguments apply to it. And in such cases, recall that shifting someone's credence enough on some issue is also morally important, even if that person is not fully convinced by one's particular argument. Philosophers and bodies of work make a larger difference collectively, and we are each responsible for the ways in which we contribute to thresholds of risks of serious harm or the prevention thereof.

Philosophers ought to pursue truth first and foremost.

There are two senses in which we might say that philosophers ought, first and foremost, to seek the truth. In the first sense, if we are doing philosophy, we ought to seek the truth first and foremost because this is a precondition for doing philosophy at all.⁵² If one shares Socrates' sentiments about the role of philosophy in Plato's *Gorgias*, one might worry that I am suggesting philosophers stop being philosophers and become mere sophists. Since my principles focus on publication and not the process of study or research, the process by which philosophers arrive at their beliefs will be very much the same as it is now. Although my principles recommend (at least as an imperfect duty) that we pursue areas of inquiry that will have the best ethical impact, it is extremely important, when researching some topic, to simply pursue truth *before* the point of publication. It would lead to morally better results if one's credence in particular viewpoints or estimations of possible moral outcomes reflect how things really are.⁵³ Of course, one might assert that if our publication practices do not respect the pursuit of truth first and foremost, we are nonetheless still not doing philosophy. This sort of assertion seems based on an overly

⁵⁰ One does not need to be Peter Singer to make a difference, either. While getting my M.A. degree, for example, I wrote a humor article based on a couple of utilitarian ideas and ended up raising over 3,000 dollars for a good charity.

⁵¹ Justin Weinberg, "Philosophy Degrees: How Many Are Awarded, And To Whom?," *Daily Nous*, February 25, 2019, <http://dailynous.com/2016/04/18/philosophy-degrees-how-many-are-awarded-and-to-whom/>.

⁵² This distinction is inspired by John Rawls, "Two Concepts of Rules," *The Philosophical Review* 64, no. 1 (1955): p. 3.

⁵³ Of course, my principles can also easily accommodate a more Rortian view of the role and nature of philosophy as well.

rigid view of what philosophy is.⁵⁴ Nevertheless, those who hold such a view may say that I advocate doing philosophy in the research phase and doing “schmilosophy” instead of philosophy during the publication phase.

There is another sense in which one might say that philosophers ought to pursue truth first and foremost. Bos van der Vossen argues that philosophers (particularly political philosophers) have a prima facie professional duty to abstain from political activism because such political engagement interferes with their task of “seeking the truth about politics.”⁵⁵ Similarly, David Benatar worries that activism in academic work risks compromising “the primary purpose of academic work,” which is “to enlighten.”⁵⁶ While the soundness of their arguments to this effect has been contested, my principles recommend a methodology of weighing the evidence as objectively as possible when doing research, so these concerns should not be a problem for my form of philosophical activism.⁵⁷

However, one might still assert that philosophers who refrain from publishing overly risky arguments, or who publish things that they have very low credence in, are failing at their professional duties to express what seems closest to the truth.⁵⁸ This may be so, but it is not unusual for moral norms to sometimes conflict with other sorts of norms, such as professional ones. We regularly face dilemmas with conflicting values or norms.⁵⁹ If I professionally ought to publish something but morally ought not to publish it, why should

⁵⁴ See, for example, Mary Midgley, “Philosophical Plumbing,” *Royal Institute of Philosophy Supplement* 33 (1992): pp. 139-151; and Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton; Oxford: Princeton University Press, 2018).

⁵⁵ Bas van der Vossen, “In Defense of the Ivory Tower: Why Philosophers Should Stay out of Politics,” *Philosophical Psychology* 28, no. 7 (2014): p. 1049. See also: Shakespeare, op. cit., p. 916-917.

⁵⁶ David Benatar, “Bioethics and Health and Human Rights: a Critical View,” *Journal of Medical Ethics* 32, no. 1 (January 2006): pp. 17-20. Cited in Draper, op. cit., p. 862.

⁵⁷ The epistemic openness I recommend here should also alleviate van der Vossen’s worry about epidemic distortion. For critiques of views like Benatar’s and van der Vossen’s, see: Rogers, op. cit., pp. 885-887; and Ben Jones, “Political Activism and Research Ethics,” *Journal of Applied Philosophy*, 2019.

⁵⁸ Perhaps philosophers could discharge their duty by pursuing truth when publishing non-risky arguments that they have high credence in, but if this is true, it only helps me by deflating the objection further.

⁵⁹ See: Derek Baker, “The Varieties of Normativity,” *The Routledge Handbook of Metaethics*, 2017, pp. 567-581; Ruth Chang, *Incommensurability, Incomparability, and Practical Reason* (Cambridge, MA: Harvard University Press, 2010).

my professional duties take precedence? After all, moral duties are usually thought to take precedence over most if not all other sorts of duties. Even if it my duty *qua* a pizza delivery driver is to deliver the pizza in thirty minutes or less, this does not trump my moral duty to drive safely. And if I have a moral duty to fulfill my professional duties, then if the usual view is correct, this duty will primarily compete with my moral duties *qua* another moral duty. If my duty to the truth were a moral duty, this would probably be because the truth-seeking role of philosophers is morally good for society. It is to this idea that we should now turn.

This argument ignores the social goods of free speech and inquiry.

Perhaps our duties to pursue the truth first and foremost win out by virtue of the morally relevant social goods they provide. For example, it might be objected that by writing and publishing articles based only on epistemic merits, we will probably discover more truths. Learning more truths is usually good for society. While only focusing on epistemic merits will sometimes cause serious harms, it will also produce many more goods. If philosophers begin withholding potentially true arguments or publishing unlikely ones in the service of moral causes, this may have bad effects overall. This rule utilitarian objection has a few controversial premises. For one thing, it is not obvious that society has usually benefited from the discovery of truths. It may be true of medicine but not true of weaponry, for example. And even if this is so, those good truths may not have been philosophical truths. It is unclear whether philosophy, or at least the kind of philosophy that my principles are concerned with, is even in the business of uncovering truths. Moral error theorists have space to argue here that ethics and applied ethics do not discover truths at all.⁶⁰ Or perhaps, as Mary Midgley suggests, the point of doing philosophy is “to turn our attention to what might be wrong in our ideas, in the structure of our thought... readjust our underlying concepts,” and “shift the set of assumptions that we have inherited and have been brought up with.”⁶¹ ‘Wrong’ here may simply mean unhelpful or *morally* wrong. Even if philosophy aims towards truths, if Bryan Frances is right, the sort of things philosophers have collectively come to know may mostly take the form of “reasons for or against certain views, elementary truths regarding fundamental notions, and highly conditionalized claims.”⁶²

⁶⁰ Many of them could say this while still endorsing my principles as first-order moral statements or useful fictions. See: Matt Lutz, “The ‘Now What’ Problem for Error Theory,” *Philosophical Studies* 171, no. 2 (2013): pp. 351-371. Talk of ‘discovering’ truths is dubious for many constructivists as well.

⁶¹ Midgley *op. cit.*, p. 140.

⁶² Bryan Frances, “Extensive Philosophical Agreement and Progress,” *Metaphilosophy* 48, no. 1-2 (2017): p. 47.

At any rate, let us meet the objection on its own assumptions. It is far from obvious that my principles would interfere with philosophy in a manner that takes away the morally relevant goods that philosophy provides. The most obvious way of testing this idea involves imagining how the history of philosophy would have played out if my principles had been used by all philosophers, and whether the goods that philosophy has contributed to—such as theoretical, scientific, cultural, political, moral, and social progress—would still be around. In the case of writing on abstract theoretical, scientific, and cultural subjects with no immediate, morally relevant practical applications, the consequences of most of such publications seem far too unforeseeable for the principles to apply. On the other hand, when it comes to writings on topics of foreseeable practical moral consequence, it is quite likely that my principles would usually have discouraged just those works that did not contribute to progress and encouraged just those works that did. For example, one might object that early feminist philosophers like Mary Wollstonecraft may have been blocked from publication for risking too much harm to society. However, the ‘harms’ to society risked by women’s suffrage is incomparable to the serious harms being done to women at the time. While the principles may have morally discouraged the publication of, for example, Kant’s defense of slavery, the world hardly seems any better because of his defense.⁶³

This suggests that an environment of epistemic inquiry free from my principles was not a necessary condition for our progress, assuming there has been such progress. This point is further supported by the fact that there never really *was* an environment of epistemic inquiry free from censorship and restrictions, so many ‘good’ truths have been learned under morally restricted conditions. Consider all the philosophers who suffered from some form of censorship or another, such as Socrates, Descartes, Spinoza, and various early social contract theorists. Of course, the censorship did not work out given that we know these philosophers, and the world may now be a better place if there had been fewer restrictions in the past. However, restrictions sometimes influence the history of thought in unexpected and unpredictable ways. For example, the Condemnations of 1210-1277 by the University of Paris against heretical works produced a skepticism against Aristotelianism that “destroyed certain essential foundations of Peripatetic physics” and inspired a skeptical turn in scholasticism from which a new natural philosophy emerged, containing seeds of early modern philosophy and science.⁶⁴

⁶³ Marx is an interesting case. Anti-Soviet Marxists will have to either say that Marx was right but should not have published, that (from the perspective of the subjective version of the negative principle) he was justified because other real or possible (and foreseeable) outcomes compensated for the risk of the disaster, or that he should have published his economic theories in a different way.

⁶⁴ Pierre Duhem, “History of Physics,” in *Catholic Encyclopedia*, ed. Charles Herbermann (New York, NY: Robert Appleton Company, 1913); David Charles Lindberg, *Science in the Middle Ages* (Chicago:

This makes it even harder to assess how exactly my principles would have changed the history of philosophy and goods provided by philosophical inquiry. The few points I make here are far from establishing how the principles really would have applied in regards to past research. To do so, we would need to dedicate hundreds of pages towards historical research. However, my responses should help to deflate our initial worry and show that those who would object to the principles on rule-utilitarian grounds share the burden of proof. Plus, even if such a historical argument could be made against me, this would not be sufficient to establish that my arguments will not be good for society from this point forth. It might be said that my principles will at least interfere with works that do good for society *sometimes*.⁶⁵ But even so, the rule utilitarian objection would only succeed if this sort of thing happened often enough to make the world a morally worse place than it would be if my principles were not used, at least in the absence of some better principles.

The objection we began with, however, could be reformulated in historically relativist terms. One might say that at any number of stages through the history of philosophy, philosophy might have been stalled by people whose values seemed to be under threat (as before, never mind that this *did* happen). It is easy to construct standards that recommend publication according to current moral knowledge, but if this had been done in the past, we would not have ended up with the moral knowledge we have now.⁶⁶ However, the subjective versions of my principles encourage epistemic humility, taking into account different possible moral theories and possibilities of serious harm according to those theories' assessments. Secondly, as Kitcher has pointed out in regards to the history of science, this objection makes controversial assumptions about the role philosophy has played as an instrument of social progress (it has often been used to defend social inequalities) and about the analogy between 'heroic liberators' in philosophy in the past and those whose arguments would be morally restricted by the negative principle today.⁶⁷ One might be worried not about the restriction of arguments but with the 'flooding' of

University of Chicago Press, 1980), p. 107-111. It is not a problem for my principles, however, that censorship sometimes has unforeseeable consequences. As I argued earlier, the only morally relevant consequences for my principles are the foreseeable ones.

⁶⁵ This is a problem for the subjective versions of my principles, since the objective versions could just endorse doing whatever leads to the least serious harm overall, no matter how things seem. The objective version is less plausible at least insofar as it is difficult to see how such a principle could be sufficiently action-guiding. See MacAskill and Ord, *op. cit.*, p. 5.

⁶⁶ I take this point analogously from an argument made about censorship and science in Kitcher, *op. cit.*, p. 298.

⁶⁷ *Ibid.*

arguments that philosophers may have low credence in. But, firstly, many untrue philosophical propositions in arguments through history have still presumably contributed to progress. Secondly, as I will later suggest, it is not particularly common or important that philosophers themselves hold a high credence in the positive principles their arguments suggest. Arguments, in some respect, cannot be ‘faked’, and my principles already contain a proviso preventing outright lies. Besides, whenever the positive principle only applies as an imperfect duty, writers are free to select cases that fulfill their professional, epistemic, and moral duties at once.⁶⁸

We should keep in mind that whenever we engage in academic inquiry and publication, we are often picking out some topics and leaving others aside not just for epistemic reasons, but partially out of considerations of social needs and goods. Out of countless subjects that we might pursue in applied ethics, we usually select the ones that seem the most morally important and pressing. Moreover, by taking certain topics like slavery off the table, we may have even *epistemically* benefitted by freeing up room to debate about more uncertain issues.

My argument, if implemented, would garner public distrust in writers and academics

Talk of ‘fake news’ and liberal bias in universities has already lessened public credulity towards institutional epistemic authorities. We would not want to make things worse, especially since academia probably deserves our trust more than institutions heavily imbued in the commercial and political sectors. If philosophers withhold information that may support certain harmful ideas, this is perfect fodder for conspiracy theories and distrust. It may even be suspected that the ‘harmful’ conclusions are more likely to be true. Alternatively, if we have an environment where potentially harmful arguments are not withheld, then people may come to trust the ‘official position’ more when it is not a harmful one.

This worry would be relevant for the application of the positive and negative principles to the sciences.⁶⁹ For the sciences, the distrust would be based on the fact that certain topics are not being empirically investigated and known empirical data is being withheld. For philosophy, however, the distrust would only be based on the fact that certain arguments that could be made were not being made. In the sciences, some theory is taken as trustworthy in light of the absence of studies that would disprove it. On the other hand,

⁶⁸ Ibid, p. 286

⁶⁹ This worry, however, is not even always valid in the case of science. See: Kitcher, op. cit. Talk of the ‘official’ vs. ‘unofficial’ position also comes from Kitcher, op. cit.

I suspect that when people are affected by philosophy, it is usually because a particular idea or argument grips them as plausible or compelling. They are usually already aware that certain arguments could and have been made against that idea, but these are arguments that the readers themselves could come up, at least in theory. No laboratory or equipment is needed; they are not relying on the philosopher in the same way that they rely on the scientist.

I also do not see many problems arising from the positive principle aside from the fact that we usually take people to be arguing for beliefs they hold high credence in. But this problem may be an insidious one. As David Foster Wallace writes, “Whether it honors them well or not, an essay’s fundamental obligations are supposed to be to the reader. The reader, on however unconscious a level, understands this, and thus tends to approach an essay with a relatively high level of openness and credulity.”⁷⁰ While I have been arguing that writers can have different obligations in different directions, we usually assume that a writer wants to tell you the truth. The reader and writer are conversing subjects, brought together in a spirit of goodwill and the pursuit of understanding. But if writers are just telling you something because they performed a calculation of the expected value of what they should tell you (whether or not they really believe it), this relationship seems broken. Wallace himself was concerned about advertisements that disguised themselves as essays. In contrast to how we approach essays, we know that “an ad’s appeal is by its very nature *calculated*—and this is part of why our state of receptivity is different, more guarded, when we get ready to read an ad.”⁷¹ Likewise, if an author is trying to convince you of something based on a calculation and not a conviction, she is not being straightforward, and there is a possibility that she is manipulating you in some way, attempting to use you as a moral instrument without your full knowledge or consent. The worry, then, is that this sort of writing offers “a perfect facsimile or simulacrum of goodwill without goodwill’s real spirit, it messes with our heads and eventually starts upping our defenses even in the cases of... true good will. It makes us feel confused and lonely and impotent and angry and scared. It causes despair.”⁷² It also risks undermining our own goals, since people would be less likely to take our arguments seriously.

For these reasons, it is important for the positive principle to have the proviso that we do not engage in outright lying or falsifying of information. However, when bearing

⁷⁰ David Foster Wallace, *A Supposedly Fun Thing I’ll Never Do Again: Essays and Argument* (Boston: Little, Brown and Co., 1997): p. 288.

⁷¹ *Ibid.*, p. 289

⁷² *Ibid.*

that proviso in mind, this potential criticism misses the point of what we are often doing when we do philosophy. As I have tried to present the positive principle, it occurs when a writer determines that the stakes in an argument are severe enough for the argument to be worth making, so long as they have sufficient credence in the argument for the expected value of the argument's consequences to be significant. This is a very roundabout way of saying that they find some argument worth considering. I can only speak from personal experience and from conversations with many people doing philosophy, but it seems common for philosophers to publish papers arguing for conclusions that they are not particularly convinced of (or even strongly doubt).⁷³ In fact, it is unusual for philosophical writings with positive proposals to assert them rather than just inviting readers to think about them.⁷⁴ Philosophers often just present an argument showing how particular premises support a particular conclusion. One usually does approach philosophy with sharpness and skepticism but not the sort with which one approaches an advertisement. There is no absence of trust or goodwill as long as the unspoken contract between the philosopher and reader remains unbroken, and the contract is just that the philosopher will present the reader with ideas that are worth considering. This contract poses no problem since an idea that one should publish according to the positive principle *is* worth considering. If moral choice-worthiness favors presenting a certain moral argument, for example, then it will also likely favor following that moral argument's recommendations. The unspoken contract between the readers and the writer, then, is unbroken. Some may think that failure to follow the belief-norm of assertion when making an argument is still a case of manipulation, but my principles can allow for this norm even if they are not tied to it.⁷⁵ Nevertheless, it is true that philosophers will have to take care to protect their individual reputations, balancing the good a particular publication might do with the possible effects it might have on that writer's credibility.⁷⁶ However, this fact should not imply that a good balance cannot be found.

Publishing harmful or risky ideas, in the end, actually has good consequences

Before concluding, there is a final objection I would like to briefly consider. As Elizabeth

⁷³ Including this paper.

⁷⁴ Thanks to Juhani Yli-Vakkuri for pointing this out to me.

⁷⁵ This issue depends on whether we should follow the belief norm of assertion and, if so, whether that norm is important enough to count as a comparable moral benefit or drawback to some decision to publish when weighed against the risk or prevention of serious harm.

⁷⁶ See Moorlock, *op. cit.*, and Rogers, *op. cit.*, p. 887.

Barnes thinks, arguments that harm can actually be a good thing. Recommending that we use a cost-benefit analysis for determining when to engage with harmful arguments, Barnes suggests that sometimes harmful arguments do nothing more than articulate preexisting but unspoken public biases and open up space for resistance against these biases.⁷⁷ On the face of things, this just means that we need to be careful about applying the negative principle, which already allows for exceptions in the cases of comparable moral benefits. However, I am skeptical that articulating unspoken prejudices will lead to their eradication more frequently than their reinforcement.

When it comes to prejudices towards underprivileged groups, Kitcher points out that there is a fragmented “moral consciousness in which public ‘politically correct’ attitudes coexist with inclinations to quite opposite beliefs” so that “we should see the impact of...research as affected both by a political asymmetry and an epistemic asymmetry.”⁷⁸ In light of this, “the history of attempts to rebut inegalitarian claims...supports the view that scientific evidence in favor of equality does not translate into further eradication of prejudices.”⁷⁹ Although philosophical arguments are probably taken as less authoritative than scientific findings, I see little reason why these general reasons would not apply to them as well. Of course, not every potentially harmful argument is one in favor of some popular prejudice, so Kitcher’s argument will only take us so far. However, many arguments that cause or risk serious harm do have this character. And while “it would be naïve to think that all pieces of controversial research are subjected to the rigors” of the ‘Millian arena,’ the negative principle will create greater epistemic demands of those publishing potentially harmful arguments.⁸⁰ These demands may sometimes do more good than the responses to harmful arguments.

VI. Conclusion

It is important to ask what the positive and negative principles imply about the place and purpose of philosophy with practical implications and in what moral direction these principles might lead philosophers. My principles remind us that when we make arguments, we always do so in a socially situated context. Our arguments in practical philosophy should take our particular audiences into account when formulating the moral

⁷⁷ Barnes acknowledges that arguments that merely defend already articulated social prejudices are not such cases.

⁷⁸ Kitcher, *op. cit.*, p. 291

⁷⁹ *Ibid.*, p. 283.

⁸⁰ Kitcher, *op. cit.*, p. 291. See also: *Ibid.*, p. 280.

recommendations that we make. For example, when advocating for particular rules to follow in situations of moral non-compliance (i.e. when other people are not doing following their moral duties), we should consider whether these very suggestions are likely to make our audiences more or less compliant. Likewise, we should consider whether the demandingness of the duties we recommend is likely to inspire or put off our audiences. We may even wish to pursue empirical research to answer these sorts of questions. A more easily fulfilled recommendation of my principles is that individual philosophers take at least take a few days to systematically consider the potential moral consequences of their work before publishing.

As for the moral implications of these principles, I suspect that they will lead us to greater altruism. They seem to promote doing philosophy in such a way that advocates for the expansion of our moral circles, erring on the side of inclusion of some group when the risk or possible prevention of serious harm is involved and would tend to advocate for more protection of marginalized and vulnerable groups. Philosophers would also probably be compelled to spend more time working on areas where serious harm is most at stake, such as poverty reduction, animal welfare, and existential risks to humanity and life. The principles may promote stronger moral duties up to a saturation point, the point at which arguing for stronger moral duties would start to exhaust and turn people away from ethics.

The conception and role of the philosopher may also change under these principles, at least where practical philosophy is concerned. The idea of a philosopher as a passive contemplator of truths may be supplanted with that of a particularly rational sort of moral activist. She would be an activist in the sense of recognizing her socially and morally embedded role as a writer.⁸¹ As such, she is also an activist in taking a normative stance and recommending certain moral views and choices (particularly, if I am right, when it comes to the expansion of our moral circles and protection of vulnerable groups). As mentioned before, she may also be an activist in the sense of becoming motivated by a possible interpretation of the positive principle to be more engaged in public dialogue on pressing moral, social, and political matters and spending more time advocating for important and morally useful philosophical positions to a general readership. She would be rational in a Pascalian sense, advocating for principles that will maximize the expected moral value of those choices, at least where serious harm is at stake (and it often is).

This is not to say that our epistemic duties will go anywhere. Although I have tried to argue that the positive and negative principles do not undermine our epistemic duties

⁸¹ See also: Hilde Lindemann, “Bioethicists to the Barricades!” *Bioethics* 33, no. 8 (2019): p. 859.

in morally important ways, some philosophers will doubtlessly prefer to evaluate their work by the traditional criteria alone, such as interestingness, originality, plausibility, etc. It is unclear to me whether we can say that loyalty to one or another of these norms is more rational or whether the choice comes down to simple preferences and partisanism.⁸² Nevertheless, those who prefer the traditional criteria should not suppose that they cancel out our moral duties and should keep in mind the possibility that good epistemology's value may primarily be instrumental in nature, serving our well-being.

Bibliography

'Average and Total Numbers of Animals Who Died to Feed Americans in 2008.' *United Poultry Concerns. Promoting the Compassionate and Respectful Treatment of Domestic Fowl*, 2009. <https://www.upc-online.org/slaughter/2008americans.html>.

Baker, Derek. 'Skepticism About Ought Simpliciter.' *Oxford Studies in Metaethics*, 13 (2018): 230–52. <https://doi.org/10.1093/oso/9780198823841.003.0011>

Baker, Derek. 'The Varieties of Normativity'. In *The Routledge Handbook of Metaethics*, edited by Tristram Colin McPherson and David Plunkett, 697. Routledge Handbooks in Philosophy. New York: Routledge, Taylor & Francis Group, 2018. <https://doi.org/10.4324/9781315213217-37>

Barnes, Elizabeth. 'Arguments That Harm - And Why We Need Them.' *The Chronicle of Higher Education*, February 2018. <https://www.chronicle.com/article/Arguments-That-Harm-and/242543>.

Benatar, David. "Bioethics and Health and Human Rights: a Critical View." *Journal of Medical Ethics* 32, no. 1 (January 2006): 17–20. <https://doi.org/10.1136/jme.2005.011775>

Bermudez, José Luis, Clare Chambers, et al. "Philosophers Should Not Be Sanctioned Over Their Positions on Sex and Gender." *Inside Higher Ed*. Accessed July 22, 2019. <https://www.insidehighered.com/views/2019/07/22/philosophers-should-not-be-sanctioned-their-positions-sex-and-gender->

Braybrooke, David. *Meeting Needs*. Princeton, NJ: Princeton University Press, 1987.

Brink, David. 'Mill's Moral and Political Philosophy'. In *Stanford Encyclopedia of Philosophy*, August 2018. <https://plato.stanford.edu/entries/mill-moral-political/#HarPri>.

Brock, Gillian, and David Miller. "Needs in Moral and Political Philosophy." *Stanford Encyclopedia of Philosophy*. Stanford University, 2019. <https://plato.stanford.edu/entries/needs/#NeedMoraObli>

⁸² See Derek Baker, "Skepticism About Ought Simpliciter," *Oxford Studies in Metaethics* 13 (2018): pp. 230-252.

- Chang, Ruth. *Incommensurability, Incomparability, and Practical Reason*. Cambridge, MA: Harvard University Press, 2010.
- Dempsey, Michelle Madden, and Matthew Lister. "Applied Political and Legal Philosophy." *A Companion to Applied Philosophy*, July 2016, 311–27. <https://doi.org/10.1002/9781118869109.ch22>
- Draper, Heather. "Activism, Bioethics and Academic Research." *Bioethics* 33, no. 8 (2019): 861–71. <https://doi.org/10.1111/bioe.12574>
- Duhem, Pierre. "History of Physics." In *Catholic Encyclopedia*, edited by Charles Herbermann. New York, NY: Robert Appleton Company, 1913.
- Feinberg, Joel. *The Moral Limits of the Criminal Law Volume One: Harm to Others*. New York: Oxford University Press, 1987. <https://doi.org/10.1093/0195046641.001.0001>
- Frances, Bryan. "Extensive Philosophical Agreement and Progress." *Metaphilosophy* 48, no. 1-2 (2017): 47–57. <https://doi.org/10.1111/meta.12227>
- Frances, Bryan. *Disagreement*. Cambridge: Polity, 2014.
- Goldberg, Sanford C. 'Should Have Known'. *Synthese* 194, no. 8 (August 2017): 2863–94.
- Hitchcock, Christopher Read. "The Role of Contrast in Causal and Explanatory Claims." *Synthese* 107, no. 3 (1996): 395–419
- Huemer, Michael. "Publishing in Philosophy," 2015. http://www.owl232.net/?fbclid=IwAR17TfXKytqSEcOemKWiigyc13Rxv8QV-UER3iy-UuUPfe2r4L_AAWb-UmI
- Jones, Ben. "Political Activism and Research Ethics." *Journal of Applied Philosophy*, 2019.
- Kagan, Shelly. *Normative Ethics*. Boulder, CO: Westview Press, 1998.
- Kitcher, Philip. "An Argument About Free Inquiry." *Noûs* 31, no. 3 (1997): 279–306. <https://doi.org/10.1111/0029-4624.00047>
- Lindberg, David. *Science in the Middle Ages*. Chicago, IL: University of Chicago Press, 1980.
- Lindemann, Hilde. "Bioethicists to the Barricades!" *Bioethics* 33, no. 8 (2019): 857–60.
- Lippert-Rasmussen, Kasper. "The Nature of Applied Philosophy." In *A Companion to Applied Philosophy*, edited by Kasper Lippert-Rasmussen, Kimberley Brownlee and David Coady. Hoboken, NJ: John Wiley and Sons, Inc., 2016. <https://doi.org/10.1002/9781118869109>
- Lockhart, Ted. *Moral Uncertainty and Its Consequences*. New York: Oxford University Press, 2000.
- Lutz, Matt. "The 'Now What' Problem for Error Theory." *Philosophical Studies* 171, no. 2 (2013): 351–71.

MacAskill, William, and Toby Ord. 'Why Maximize Expected Choice-Worthiness?' *Noûs*, 14 July 2018. <https://doi.org/10.1111/nous.12264>

Midgley, Mary. 'Philosophical Plumbing'. *Royal Institute of Philosophy Supplement* 33 (September 1992): 139–51.

Mill, J. S. *On Liberty*. London: Longmans, Green, Reader and Dyer, 1869. https://en.wikisource.org/wiki/On_Liberty.

Mohr, Noam. Average and Total Numbers of Animals Who Died to Feed Americans in 2008. United Poultry Concerns, August 2009. <https://www.upc-online.org/slaughter/2008americans.html>.

Moorlock, Greg. "Do Junior Academic Bioethicists Have an Obligation to Be Activists?" *Bioethics* 33, no. 8 (2019): 922–30. <https://doi.org/10.1111/bioe.12649>

Pierre Duhem. 'History of Physics'. In *The Catholic Encyclopedia*, edited by Charles George Herbermann. New York: Robert Appleton Company, 1913.

Pogge, Thomas. *World Poverty and Human Rights*. Cambridge: Polity, 2008.

Rawls, John. "Two Concepts of Rules." *The Philosophical Review* 64, no. 1 (1955): 3.

Rogers, Wendy. "Bioethics and Activism: A Natural Fit?" *Bioethics* 33, no. 8 (August, 2019): 881–89. <https://doi.org/10.1111/bioe.12558>

Rorty, Richard. *Philosophy and the Mirror of Nature*. Princeton; Oxford: Princeton University Press, 2018.

Rowbottom, Darrell P. *Probability*. Cambridge, UK: Polity, 2015.

Royzman, Edward, and Jonathan Baron. "The Preference for Indirect Harm." *Social Justice Research* 15, no. 2 (2002). <https://www.sas.upenn.edu/~baron/papers/indirect.pdf>.

Schaffer, Jonathan. "The Metaphysics of Causation." Stanford Encyclopedia of Philosophy. Stanford University, July 5, 2016. <https://plato.stanford.edu/entries/causation-metaphysics/>.

Scheffler, Samuel. "Ethics Discussion at PEA Soup: Matthew Rendall's 'Discounting, Climate Change and the Ecological Fallacy,' with a Critical Précis by Samuel Scheffler." Edited by Chike Jeffers. PEA Soup. Prindle Institute for Ethics, April 14, 2019. <http://peasoup.us/2019/04/ethics-discussion-at-pea-soup-matthew-rendalls-discounting-climate-change-and-the-ecological-fallacy-with-a-critical-precis-by-samuel-scheffler/>.

Scully, Jackie Leach. "The Responsibilities of the Engaged Bioethicist: Scholar, Advocate, Activist." *Bioethics* 33, no. 8 (2019): 872–80. <https://doi.org/10.1111/bioe.12659>

Shakespeare, Tom. "When the Political Becomes Personal: Reflecting on Disability Bioethics." *Bioethics* 33, no. 8 (2019): 914–21. <https://doi.org/10.1111/bioe.12668>

Toppinen, Teemu. “The Ethics of Dietary Choice: Does it Matter?” *The Philosophy of Food*. Class lecture at the University of Helsinki, Helsinki, Finland, May 4, 2018).

Unger, Peter K. *Living High and Letting Die: Our Illusion of Innocence*. New York, NY: Oxford Univ. Press, 2011. <https://doi.org/10.1111/0029-4624.00093>

Van Mill, David. ‘Freedom of Speech’. In *Stanford Encyclopedia of Philosophy*, 2017. <https://plato.stanford.edu/entries/freedom-speech/#ResHarPri>.

Vossen, Bas van der. ‘In Defense of the Ivory Tower: Why Philosophers Should Stay out of Politics.’ *Philosophical Psychology* 28, no. 7 (3 October 2015): 1045–63. <https://doi.org/10.1080/09515089.2014.972353>

Waldron, Jeremy. *The Harm in Hate Speech*. First Harvard University Press paperback edition. Cambridge, Massachusetts London, England: Harvard University Press, 2014. <https://doi.org/10.21875/tjc.v3i1.22486>

Wallace, David Foster. *A Supposedly Fun Thing I'll Never Do Again: Essays and Arguments*. 1st ed. Boston: Little, Brown and Co, 1997. <https://doi.org/10.1176/appi.ajp.2011.11060977>

Weinberg, Justin. “Philosophy Degrees: How Many Are Awarded, And To Whom?” *Daily Nous*, February 25, 2019. <http://dailynous.com/2016/04/18/philosophy-degrees-how-many-are-awarded-and-to-whom/>.

Weinberg, Justin. “Recognizing Gender Critical Feminism as Anti-Trans Activism (Guest Post).” *Daily Nous*, August 6, 2019. <http://dailynous.com/2019/08/06/recognizing-gender-critical-feminism-anti-trans-activism-guest-post/>.

“Why Nets?” *The Against Malaria Foundation*. Accessed October 20, 2019. <https://www.againstmalaria.com/WhyNets.aspx>.

Williams, Bernard. “Morality, the Peculiar Institution.” In *Ethics and the Limits of Philosophy*. London: Routledge, 2011.

Wolf, Susan. “Moral Saints.” *The Journal of Philosophy* 79, no. 8 (1982): 419.

Zion, Deborah. “On Beginning with Justice: Bioethics, Advocacy and the Rights of Asylum Seekers.” *Bioethics* 33, no. 8 (2019): 890–95. <https://doi.org/10.1111/bioe.12660>