Philosophical Methodology: A Plea for Tolerance*

Sam Baron[1], Finnur Dellsén[2], Tina Firing[3], and James Norton[4]

[1]University of Melbourne; s.baron@unimelb.edu.au

[2]University of Iceland, Inland Norway University of Applied Sciences, & University of Oslo; fud@hi.is

[2]University of Iceland; tinafiring@hi.is

[4]University of Tasmania; james.norton@utas.edu.au

Abstract

Many prominent critiques of philosophical methods proceed by suggesting that some method is *unreliable*, especially in comparison to some alternative method. In light of this, it may seem natural to conclude that these (comparatively) unreliable methods should be abandoned. Drawing upon work on the division of cognitive labour in science, we argue things are not so straightforward. Rather, whether an unreliable method should be abandoned depends heavily on the crucial question of how we should divide philosophers' time and effort between different methods, in order to maximise our prospects of achieving epistemic success. We show that, in a range of cases, even a (comparatively) unreliable method deserves to be allocated some of these resources.

1 Introduction

Philosophical methods are under attack. From analytic metaphysics (Ladyman and Ross, 2007), to the method of cases (Machery, 2017), to experimental philosophy (Cappelen, 2012), debates about philosophical methods are replete with philosophers arguing against one method or another. Arguments of this kind, where successful, tend to support the

*This article is forthcoming in *Analysis*. Please cite the published version when available.

idea that the targeted philosophical method is unreliable. This gives rise to an important question: should unreliable philosophical methods be abandoned?

Drawing upon work on the division of cognitive labour in science, we argue that in order to adjudicate whether and to what extent a given method should be abandoned, we must address the broader question of how we should divide our limited resources—i.e., philosophers' time and effort!—between different methods, in order to maximise our prospects of achieving epistemic success. We show that, sometimes, even relatively unreliable methods should be given their due, by receiving some allocation of resources. This is significant, we suggest, as it provides a clear way forward in arguing for methodological pluralism.[1]

2 Unreliability Critiques of Philosophical Methods

Philosophers often criticise methods used within the discipline. Consider, for example, Ladyman and Ross's (2007) pronouncement that what they variously call 'analytic', 'a priori', or 'neo-Scholastic' metaphysics "should be discontinued" (2007, vii). Ladyman and Ross do not merely claim that their preferred brand of more scientifically engaged metaphysics is a *better* way to investigate metaphysical questions. Instead, they propose the complete cessation of traditional analytic metaphysics.[2]

As another example, consider the 'negative program' within experimental philosophy. As the movement's unofficial logo—the burning armchair—suggests, this program has generated influential critiques of *a priori* methods. For instance, Weinberg (2007, 327) claims that "intuition—or at least our current philosophical practice involving it—is hopeless." Similarly, Machery (2017, 148) proposes that "except when cases are known to elicit a broad consensus outside philosophy, philosophers should shelve the method of cases."

Various methods employed by experimental philosophers have been targeted in turn.

[1]Calls for methodological pluralism can be detected in, e.g., Andow (2021, 2022); Knobe and Nichols (2007); Kornblith (2013); Strevens (2019); Williamson (2021).

[2]See also Carnap (1959); French and McKenzie (2012); Maclaurin and Dyke (2010); McKenzie (2020).

Cappelen (2012, 222), for instance, claims that "[t]he project of checking on people's intuitions is philosophically pointless" and that experimental philosophy should be left behind, along with "methodological rationalism" (2012, 19). Cappelen's more recent book *Fixing Language* attacks a much broader target. In the subsection aptly entitled: "Implications for Philosophical Methodology: Purely Descriptive Philosophy Must Be Abandoned", Cappelen argues that ways of doing philosophy that don't involve conceptual engineering should be left behind (2018, 47).

There are, broadly, at least two ways in which a method may be flawed. First, a method may be such that it has zero chance of achieving the desired result. Such a method cannot possibly succeed. By contrast, a method may be such that it can possibly succeed, but is nonetheless an unreliable means of achieving success, either in absolute terms or in comparison to some alternative method.

A method would have to be very badly flawed in order for its chances of success to be zero. For example, suppose a method is *self-defeating*, in the sense that the method itself undermines any possibility of its own success. Or suppose that a method is *internally contradictory*, such that it cannot be used coherently to achieve anything. We rarely see arguments pointing out such decisive flaws. What we see, instead, are criticisms of methods that point more in the direction of unreliability.

Ladyman and Ross, for instance, argue that analytic metaphysics relies too heavily on accordance with common sense as a method for discerning objective truths about reality. This is a problem, they argue, because 'common sense' is a poor basis upon which to discern such truths, since it is infected by everyday prejudices and misunderstandings that have a distorting effect. Notice, however, that if someone is lucky enough to have the right common sense intuitions, then relying on such intuitions may well be a route to objective truths. So the chance of success is not zero. Still, if Ladyman and Ross are correct, the chance of success is very low: reliance on common sense is an unreliable strategy for metaphysics.

A similar point holds for the experimental-philosophy-based critique of using intuitions as evidence. If we grant that most people's intuitions are parochial and fallible—perhaps disastrously so—it follows that this method is unreliable. However, it does not follow that its chance of getting things right is zero. For again, one might simply get lucky and happen to have the right (albeit culturally influenced) intuitions. Still, the criticism suggests that using intuitions is unreliable, and perhaps indeed it is.[3]

Our concern in this paper is with what follows from a (successful) argument that a method is unreliable. Should the method in question be abandoned? It is very tempting to think so, and this inclination is reflected in some of the works discussed above. Why use an unreliable method, especially if a more reliable alternative method is available? Given a choice of methods with which to build a house, if one is unlikely to succeed, while another is not, it seems that one should abandon the method with poor prospects. As we will argue, however, things are not so straightforward when it comes to philosophical methodology. Doing philosophy is not much like building a house — sometimes it pays to have someone using an unreliable method.[4]

3 Methodological Division of Cognitive Labour

An isolated epistemic agent, seeking to improve their cognitive state with regard to some phenomenon, has meagre resources upon which to draw. It's simply not feasible for them to try out all the available methods simultaneously and see what comes of each. It will thus normally be rational for such an agent to exclusively adopt the method that appears to be the most promising. However, things are rather more complicated for the collective enterprise of philosophy. For when a philosopher in the collective is choosing

[3]To be clear, we are not saying that the only way to criticise a method is to show that it is unreliable. Nor are we saying that this is all the above arguments show. The point is just that arguments about methods often provide reasons to take at least some methods in philosophy to be (comparatively) unreliable. Moreover, this is to be expected: every field has a variety of methods, some of which are more reliable than others, and some of which may be quite unreliable.

[4]Andow (2021) argues for a somewhat similar conclusion regarding conceptual engineering specifically. Andow's argument is, roughly, that the potential societal benefits of conceptual engineering, if successful, may be sufficiently great to justify the costs of engaging in it, even if the probability of success is very low. As we'll see, our model below does not involve this sort of cost-benefit analysis, but focuses rather on the division of cognitive labour amongst philosophers.

between methods for investigating some phenomenon, they do so against a background of information about which methods other philosophers are using to investigate the same phenomenon.

In an influential paper, Kitcher (1990) calls the matter of how a community of researchers should allocate their resources 'the division of cognitive labour'. Kitcher's main focus is on how the scientific enterprise incentivises researchers to distribute their efforts between methods so as to maximise the probability of the community achieving epistemic success, rather than having each researcher adopt the method that appears to be the most promising. What is most important for us, however, is the simple Peircean (1958) point that the *optimal distribution*—the distribution of individual researchers between various methods which maximises the chance of the community achieving epistemic success—will in some circumstances require that some researchers adopt methods other than the one that appears to be the most promising.

This point is most clearly illustrated using an extension of the model of resource allocation articulated by Strevens (2003) (which in turn is influenced by Kitcher (1990)). First, let's call the application of a given method to investigate some phenomenon a *program*, and follow Strevens in making the idealising assumption that a program can either totally succeed in its investigation of its target phenomenon, or totally fail. Likewise, we will follow Kitcher in assuming that each program has a distinct *intrinsic potential* which determines its chance of success given the investment of some specific amount of resources. With these assumptions in place, we can ascribe to each program a *success function*, s_i, which maps any amount of resources invested in it and competing programs to a likelihood of success and failure of these programs.[5]

In what follows, we will be comparing various distributions of resources between these two hypothetical programs—Program 1 and Program 2—with an eye towards determining

[5]Specifically, in the case of two programs, $s_1(n_1, n_2)$ is the likelihood of success in Program 1 and failure in Program 2 given that n_1 resources are invested in Program 1 and n_2 are invested in Program 2. Similarly, $s_2(n_1, n_2)$ is the likelihood of success in Program 2 and failure in Program 1, given the same investment of resources. As explained in footnote 6, these functions reduce to $s_1(n_1)$ and $s_2(n_2)$ respectively, given the simplifying assumptions made below.

what the optimal distribution would be under various assumptions about their success functions s_1 and s_2. We will make three simplifying assumptions primarily to make the model easier to work with. These assumptions can be removed, and the model de-idealised, at the price of increased mathematical complexity. The first assumption is that the utility of success arising from either of the two programs is equal, since these are competing programs aiming to investigate the very same phenomenon. Formally, $v_1 = v_2$, where v_1 and v_2 denote the values of success in each program. The second simplifying assumption is that the amount of resources invested into one program does not affect the likelihood of success in another program. This implies that a program's likelihood of success is simply a function of the amount of resources invested into that program, as opposed to also being a function of the resources invested into a competing program.

The third simplifying assumption we make is that the two programs cannot both succeed (although they might both fail). Put differently, the assumption is that the two programs are *mutually exclusive* with respect to their success: if one succeeds, the other fails. To be clear, we do not think this is a particularly plausible assumption to make about typical comparisons of different philosophical programs, since it often seems perfectly possible for both to be successful. For example, even if naturalised metaphysics à la Ladyman and Ross achieves success regarding some particular phenomenon, that does not automatically imply that the type of *a priori* metaphysics they criticise cannot *also* achieve success regarding that very same phenomenon.

However, the implausibility of this simplifying assumption does not undermine the conclusions we draw below, since those conclusions would only be strengthened if it were instead assumed that both programs could succeed. If so, there would be two types of situation in which devoting some resources to Program 1 and also to a supposedly less promising Program 2 would pay off: (i) situations in which Program 2 succeeds while Program 1 fails, and (ii) situations in which both programs succeed. Our simplifying assumption essentially eliminates the possibility of (ii), thus leaving (i) as the only type of situation in which devoting resources to Program 2 would actually pay off. Since our overall concern is precisely to argue that devoting resources to Program 2 is often optimal from the point of view of expected returns, this stacks the deck against the position we will be arguing for. Even with the deck thus unfavorably stacked, we shall see that the optimal distribution of resources will often assign some resources to the supposedly less promising Program 2.

With these simplifying assumptions in place, we can construct a notion of *expected return* (ER) by analogy with the notion of expected utility. The expected return on investing some amount of resources in different programs is the sum of the products of the utilities and probabilities of all the different combinations of success and failure given these investments. In the case of only two programs, and with our simplifying assumptions in place, this reduces to:

$$ER(n_1, n_2) = v\big(s_1(n_1) + s_2(n_2)\big) \tag{†}$$

where v is the value of succeeding in either program, and $s_1(n_1)$ and $s_2(n_2)$ are the probabilities of success of Program 1 and Program 2, respectively, given that n_1 and n_2 resources are invested in each.[6]

Given this, one can work out the optimal distribution of resources, i.e., the distribution

[6] The more general formula from which (†) is derived states that the expected return is the sum of the products of the respective values of success and probabilities for (i) succeeding in Program 1 while failing in Program 2, (ii) failing in Program 1 while succeeding in Program 2, and (iii) succeeding in both Program 1 and Program 2 (we assume failing in both programs is an outcome with no value):

$$ER(n_1, n_2) = v_1 s_1(n_1, n_2) + v_2 s_2(n_1, n_2) + v_{1\&2} s_{1\&2}(n_1, n_2) \tag{‡}$$

(‡) reduces to (†) given our simplifying assumptions via the following four steps: First, the third assumption implies that $s_{1\&2}(n_1, n_2) = 0$, eliminating the third term. Second, that same assumption also implies that the probability of success in Program 1 and failure in Program 2 just is the probability of success in Program 1, since success in Program 1 would entail failure in Program 2. Now, by the second assumption, this probability of success in Program 1 is not affected by the number of resources invested in Program 2, so $s_1(n_1, n_2)$ can be written more simply as a function of n_1 alone: $s_1(n_1)$. By parity of reasoning, $s_2(n_1, n_2)$ also simplifies to $s_2(n_2)$. Third, the first assumption is that $v_1 = v_2$, so both terms can be written as v. Finally, factoring this v out from the remaining two terms above gets us (†).

that will maximise expected return across both programs. Assuming there is a fixed total of $N = n_1 + n_2$ resources to distribute between the programs, this becomes a simple optimisation problem regarding the distribution of N resources between the first program (n_1) and the second (n_2) which makes it most likely that one of the two programs delivers success. How such an optimisation problem is solved in a given case will clearly depend on the specific properties of the success functions s_1 and s_2, so there is no universally correct way of dividing up N resources between competing programs. Despite this, it is still possible to discern several general implications of our model that speak to the issue of how philosophers ought to distribute their resources across different programs.

Of particular relevance are various circumstances in which expected returns will not be maximised by devoting all resources to a single program, i.e., that the optimal distribution won't be $N = n_1$ or $N = n_2$. One might think that this will only happen when there is something degenerate about one program's success function. For example, suppose that Program 1 offers substantial expected returns for any investment of resources up to some amount n_t, but that investing resources beyond n_t *decreases* the program's probability of success. Given the option of investing in a second program, Program 2, with a success function that increases monotonically but initially more slowly, one should clearly devote at most n_t resources to Program 1 and devote the remaining resources to Program 2 (Figure 1). In this type of case, Program 1 might be one in which 'too many cooks spoil the broth', perhaps because once the community of researchers reaches a certain size they start getting in one another's way.

Importantly, (‡) explains why the second simplifying assumption stacks the deck against the position for which we'll be arguing, i.e., that investing some resources in both programs is often optimal. By eliminating the third term, this assumption takes away a potential benefit to investing some resources into Program 2 in addition to Program 1, viz. that by doing so one ensures that this term is not reduced to zero. Indeed, it is worth noting that this term might contribute significantly to the expected return even when the probability of both programs succeeding is low, because the value of this eventuality is arguably often quite high, e.g., in that it allows for *methodological triangulation* (Heesen et al., 2019; Trpin, 2023).

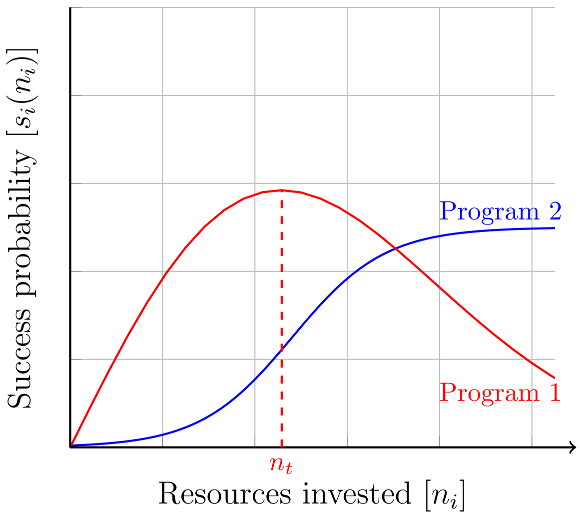

Figure 1: Program 1 becomes 'degenerate' after n_t resources.

A less obvious type of situation in which neither $N = n_1$ nor $N = n_2$ maximises expected returns is one in which (i) Program 1 requires very little resources to increase its chances of success to some extent, but adding further resources to the program doesn't increase its chances further, and (ii) Program 2 requires much more substantial resources to increase its chances of success at all, but once these resources are devoted to the program, its chances of success are much higher than the first program's (Figure 2). If N is small, it is optimal to devote all resources to Program 1. However, if N is sufficiently large, it is optimal to devote a smaller amount of resources to Program 1 and a larger amount to Program 2.

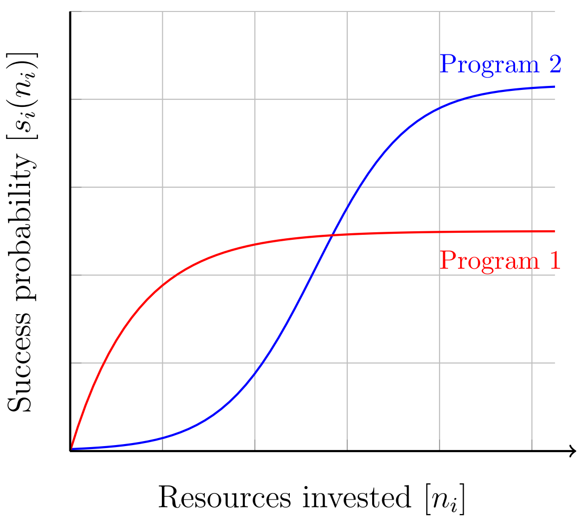

Figure 2: Program 1 is 'better' for meagre resources; conversely for Program 2.

This example also highlights that although it is tempting to think of programs as the sort of thing that can simply be ranked from 'better' to 'worse' in terms of their intrinsic propensities for delivering success, this can also be highly misleading. A program's chance of success depends not only on its intrinsic propensity, but also on the amount of resources devoted to it. In Figure 2, Program 1 is 'better' for a small amount of resources, in that it delivers some substantial chance of success at a low resource cost; while Program 2 is 'better' for larger amounts of resources.

With that said, there can also be programs whose chances of success are better than some other program's for any given amount of resources devoted to each. That is, for some programs' success functions s_1 and s_2, it will be the case that, for any $n \in \mathbb{R}_{>0}$, $s_1(n) > s_2(n)$. In that case, let us say that the first program *strictly dominates* the second. Now, there is a clear sense in which a program that strictly dominates another is thereby better than the dominated program. Nevertheless, it does not follow that the optimal distribution of resources will be such that the dominating program receives all available resources.

The reason for this is quite simple. At some point, devoting additional resources to the dominated program might contribute more to the overall expected return than would be contributed by devoting the same amount of additional resources to the dominating program. This will be clearest in cases in which both programs have diminishing marginal expected returns, i.e., are such that additional resources beyond a certain point contribute less and less to the overall expected return (see Figure 3). In that case, devoting even more resources to a program to which a lot of resources are already devoted will contribute less to the expected return than devoting those same resources to a program to which much fewer resources are currently being devoted, even if the first strictly dominates the second, and indeed even if the second program utilises an unreliable method.

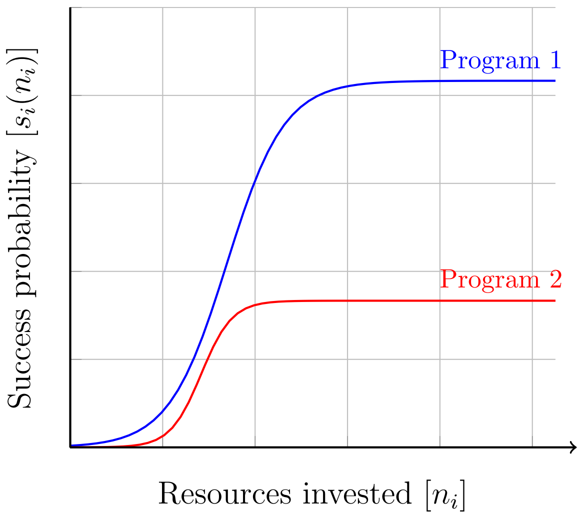

Figure 3: Program 1 strictly dominates Program 2.

4 Upshots for Debates about Philosophical Methods

Thinking about the division of cognitive labour reveals that a collective epistemic enterprise will not always maximise its chances of achieving its epistemic goals by having every researcher adopting the same method—even if that method strictly dominates all other available methods. With this insight in hand, we turn back to the question of whether and to what extent an unreliable philosophical method should be abandoned.

In the first instance, it is useful to distinguish two kinds of abandonment: abandonment of a method forever—abandonment *in aeternum*—and abandonment of a method for the time being—abandonment *pro tem*. Our discussion reveals that whether a method should be abandoned in either sense depends heavily on three things. First, on the resource landscape: the current resources available for distribution. Second, on the shape of the method's success function: the function from resources to probability of success. And third, on comparative information about *other methods'* success functions.

Given this, it is very hard to see how one could argue from the unreliability of a method to its abandonment *in aeternum*. In order to show that a method should be abandoned forever, one would need to show that no matter what resources we might have available in the future, and no matter if all competing programs are plateauing, it will never be optimal to invest any resources into that method. We seriously doubt anyone's capacity to show any such thing. Indeed, it is not clear that one can speak meaningfully about the reliability of a method without specifying some level of resource investment, since as we've seen in §3, a method can have a very low probability of success with a small resource allocation, and yet have a very high probability of success given additional resources.

What about abandonment *pro tem*? Well, it simply does not follow from the fact that a given method is unreliable that it ought to be abandoned *pro tem*. No amount of analysing the effectiveness of a method in isolation from facts about resources and other methods will be sufficient for making an argument of this kind. Rather, we must take into account the bigger picture, considering other methods and their prospects as well as the resources available. If we do that, it is indeed possible to argue for the abandonment of a method *pro tem*. Accordingly, it may be that some of methods discussed in §2 really should be set aside for now, but this is by no means guaranteed by arguments seeking to demonstrate their unreliability.

Parallel reasoning applies to arguments *in favour* of a given method. Arguing that a certain method *should be used* requires the same attention to the three features we have highlighted: resource landscape, shape of the method's success function, and comparison with other methods' success functions. Again, even methods that boast substantial virtues may not warrant the allocation of more resources, for instance if they've already hit the point of significantly diminishing returns (and some other method hasn't).

All of this points in the direction of methodological pluralism in so far as it speaks against a winner-takes-all approach in which all resources are allocated to the most reliable method. That being said, it does not follow from the framework that we have offered that methodological pluralism is appropriate in the investigation of all philosophical phenomena at all times, e.g., because there will be cases in which the limited resources we have available are only sufficient to generate one critical mass of the sort required for any significant return on the invested resources. However, as we have seen, there are multiple other situations, including situations in which one method strictly dominates another, in which our resources—especially if somewhat less limited—should be allocated to a plurality of different programs, each using a different method.

To be clear, determining the prospects of various methods is no easy task; our vision of these debates does not provide a shortcut to easy answers to difficult questions about the available methods' success functions and how they compare to each other. Still, it is important to have a clear view of *which* difficult questions we must tackle in order to make sensible recommendations about philosophical methodology.[7]

References

Andow, J. (2021). Conceptual engineering is extremely unlikely to work. so what? *Inquiry 64*(1), 212–226.

Andow, J. (2022). How to vindicate the armchair. *Analysis 82*(2), 306–321.

Cappelen, H. (2012). *Philosophy Without Intuitions.* Oxford: Oxford University Press.

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering.* Oxford University Press.

Carnap, R. (1959). The elimination of metaphysics through logical analysis of language. In A. J. Ayer (Ed.), *Logical Positivism*, pp. 60–81. The Free Press.

French, S. and K. McKenzie (2012). Thinking outside the toolbox: Towards a more productive engagement between metaphysics and philosophy of physics. *European Journal of Analytic Philosophy 8*(1), 42–59.

Heesen, R., L. K. Bright, and A. Zucker (2019). Vindicating methodological triangulation. *Synthese 196*(8), 3067–3081.

Kitcher, P. (1990). The division of cognitive labor. *Journal of Philosophy 87*(1), 5–22.

Knobe, J. and S. Nichols (2007). An experimental philosophy manifesto. In J. Knobe and S. Nichols (Eds.), *Experimental Philosophy.* Oxford University Press.

[7]For valuable feedback on earlier versions of this paper, we are grateful to Jessica Pohlmann, audience members at the Eastern Hemisphere Language and Metaphysics Network and at the University of Bergen, and two anonymous referees for this journal.

Kornblith, H. (2013). 12 is there room for armchair theorizing in epistemology? *Philosophical methodology: The armchair or the laboratory?*, 195.

Ladyman, J., D. Ross, D. Spurrett, and J. Collier (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Machery, E. (2017). *Philosophy Within Its Proper Bounds*. Oxford: Oxford University Press.

Maclaurin, J. and H. Dyke (2010). What is analytic metaphysics for? *Australasian Journal of Philosophy 90*(2), 291–306.

McKenzie, K. (2020). A curse on both houses: Naturalistic versus a priori metaphysics and the problem of progress. *Res Philosophica 97*(1), 1–29.

Peirce, C. S. (1958). *The Collected Papers of Charles Sanders Peirce*. Harvard University Press.

Strevens, M. (2003). The role of the priority rule in science. *Journal of Philosophy 100*(2), 55–79.

Strevens, M. (2019). *Thinking Off Your Feet: How Empirical Psychology Vindicates Armchair Philosophy*. Oxford University Press.

Trpin, B. (2023). Against methodological gambling. *Erkenntnis 88*(3), 907–927.

Weinberg, J. M. (2007). How to challenge intuitions empirically without risking skepticisim. *Midwest Studies in Philosophy 31*(1), 318–343.

Williamson, T. (2021). *The philosophy of philosophy*. John Wiley & Sons.