Filozofia Nauki (*The Philosophy of Science*)
ISSN 1230-6894 e-ISSN 2657-5868
2021, vol. 29(2) [114]: 77-113
DOI: 10.14394/filnau.2021.0012

ROBERT BARNARD,* JOSEPH ULATOWSKI,** JONATHAN M. WEINBERG***

THE FOURFOLD ROUTE TO EMPIRICAL ENLIGHTENMENT
EXPERIMENTAL PHILOSOPHY'S ADOLESCENCE AND THE CHANGING BODY OF WORK****

Abstract

The time has come to consider whether experimental philosophy's ("x-phi") early arguments, debates, and conceptual frameworks, that may have worn well in its early days, fit with the diverse range of projects undertaken by experimental philosophers. Our aim is to propose a novel taxonomy for x-phi that identifies four paths from empirical findings to philosophical consequences, which we call the "fourfold route." We show how this taxonomy can be fruitfully applied even at what one might have taken to be the furthest edges of possible applications of x-phi in metaphysics and formal philosophy. Ultimately, the fourfold route helps us understand a different kind of empirical fact: the development of x-phi itself.

Keywords: metaphilosophy, experimental philosophy, methodology, history of analytic philosophy, empirical philosophy, ordinary language philosophy, Arne Næss, J. L. Austin

* University of Mississippi, Department of Philosophy and Religion, 100 Bryant Hall, University, Mississippi 38677 USA, e-mail: rwbjr@olemiss.edu, ORCID: https://orcid.org/0000-0002-1702-9016.

** University of Waikato, Philosophy Programme, School of Social Sciences, Division of Arts, Law, Psychology, and Social Sciences, Private Bag 3105, Gate 1, Knighton Road, Hamilton 3240 New Zealand, e-mail: joe.ulatowski@waikato.ac.nz, ORCID: https://orcid.org/0000-0003-3019-1269.

*** University of Arizona, Department of Philosophy, Social Sciences, Room 213, 1145 E. South Campus Drive, Tucson, Arizona 85721 USA, e-mail: jmweinberg@email.arizona.edu, ORCID: https://orcid.org/0000-0001-9075-4727.

**** We are grateful to the special-issue editor, Adrian Ziółkowski, Manuel Isaac Gustavo, Liezl van Zyl, as well as two very thorough referees for their constructive feedback on our paper.

> But if someone were to say "So logic too is an empirical science" he would be wrong. Yet this is right: the same proposition may get treated at one time as something to test by experience, at another as a rule of testing.
>
> Ludwig Wittgenstein, *On Certainty*, § 98

Questions about philosophical methodology have come into explicit focus in recent years. In this context, the rise of experimental philosophy ("x-phi") has occasioned an ongoing discussion about the scope and place of x-phi as one philosophical approach among others. Assessing experimental philosophy now, in its late teenage years, the time has come to consider whether some of its early arguments, debates, and conceptual frameworks, which may have worn well in its early days, fit with the diverse range of projects undertaken by experimental philosophers. We seek here to take some metaphilosophical measurements on both a larger and smaller scale. From a zoomed-out vantage point, we discard a metaphilosophical accusation that is both completely outgrown and thoroughly out of style, which we call metaphilosophical majoritarianism; and we tailor and amend a conceptual tool that goes back to x-phi's infancy, the distinction between the positive and negative program. Our more fine-grained focus will be on how these issues play out at what one might have taken to be the furthest edges of possible applications of x-phi — namely, to metaphysics and formal philosophy.

1. EXPERIMENTAL PHILOSOPHY, A COMING-OF-AGE STORY

It seems to us that there is not much of a useful debate to be had anymore as to whether x-phi counts legitimately as philosophy, or as to whether it frequently (though perhaps not always) meets the standards of proper work in the social sciences. Still up for debate, however, is what role (or roles) that x-phi might play in philosophical work. Going forward, should x-phi be restricted to merely showing the error of our intuitive ways, or just uncover psychological facts that are of intrinsic philosophical interest? Might x-phi help service the intuitive beginnings of rich, philosophical arguments?

Practitioners of x-phi have often had to protest against a rather narrow and blinkered misconception of what x-phi is and how it is meant to be brought to bear on philosophical questions. As one observer reports (without approving of it):

> It's very common to see professional philosophers roll their eyes at the mere mention of x-phi. There are probably numerous reasons for the eye rolling, but one simple thought process here goes something like this:
>
> *Eye-rolling response to x-phi*: The idea that we can determine the answers to substantive philosophical questions by asking ordinary folk what they think is just ludicrous. It's analogous to claiming that we can answer open questions in physics or mathematics by polling ordinary folk. Who cares what the folk think? (Balaguer 2016: 2368)

This view sees x-phi as subjecting philosophical theses to a kind of opinion polling, according to a rather brute inference *ad populum*. When one adopts this view of x-phi, it is relatively easy to dismiss its philosophical relevance. Other critiques of x-phi have offered a disguised version of this objection by suggesting that the methods and experimental designs used in x-phi studies are naïve and unsophisticated (cf. Cullen 2010, Horvath 2010, Woolfolk 2013). When a critic talks about x-phi in terms of a "poll," it often means that they are going to be dismissive of it.[1] Let us call the view that x-phi amounts to opinion polling *ad populum* "metaphilosophical majoritarianism" (MM for short).

While it should be clear from reading a handful of contributions to the literature that MM is a gross mischaracterization of what experimental philosophers are up to, nonetheless, there really is a legitimate question of just how to get from psychological data — that is, x-phi results in general — to the kinds of substantial first-order philosophical results about matters like the nature of truth, knowledge, or justice.[2] There is, in general, an apparent inferential gulf between the substantive philosophical debates and the empirical results put forward as relevant to them. On one side of that gulf, analytically trained philosophers make arguments about compatibilism or incompatibilism (see, e.g., McKenna, Pereboom 2014), the priority of knowledge over belief (e.g., Williamson 2000), or the proper solution to the Liar Paradox (e.g., Jago

---

[1] However, Frank Jackson (1998), in a bit of discussion of x-phi *avant la lettre*, explicitly advocates opinion polls as part of the Canberra Program! He writes: "I am sometimes asked . . . why, if conceptual analysis is concerned to elucidate what governs our classificatory practice, don't I advocate doing serious opinion polls on people's responses to various cases? My answer is that I do — when it is necessary" (Jackson 1998: 36-37). Though, unfortunately, he also claims that a professor asking students in class is an adequate methodology. And he says that "often we know that our own case is typical and so can generalize from it to others" (Jackson 1998: 37). To summarize Jackson's position: if our own case is typical, then we need not query ordinary people about their own intuitions. If we do query others, our students are enough.

[2] Here are some examples of good experimental work being done by some local philosophers who show how x-phi is not just a matter of polling non-philosophers: Żuradzki, Wiśniowska 2020 and Paprzycka-Hausman 2015.

2018). On the other side, we have information about how different stimuli or experimental circumstances elicit different patterns of responses from groups of participants (e.g., Nadelhoffer et al. 2004, Rose 2015, Barnard, Ulatowski, Weinberg 2017). How on earth — or even off earth — are we to get from one side to the other? MM assumes that x-phi practitioners mean to bridge that distance by simply denying that there ever was a gulf in the first place.[3] While that obviously will not work as a general solution, it usefully raises the question: just how *can* we build a bridge from empirical facts to philosophical truths? What strategies are available here to make x-phi relevant to traditional philosophical questions?

Philosophers familiar with the traditional distinction between the "negative program" and "positive program" of x-phi may wonder whether that distinction already encapsulates the approaches to those questions. To begin, let us introduce the positive/negative program distinction with a widely accepted description of it:

> Some experimental philosophers [in the positive program] are quite philosophically conservative: they see their work as aiding in the traditional philosophical project of conceptual analysis. Others [in the negative program] see their work as primarily disruptive: they argue that empirical research into the nature and sources of our intuitions reveals that they are ill suited to serve as the foundation for philosophical theorizing. (Plakias 2015: 6)

Our proposed taxonomy will crosscut this distinction and will ultimately supplement it in two ways. First, the negative/positive *distinction* can be mistaken for a *partition* between work that is meant only to undermine the evidential status of intuitions and work that is meant only to provide support for a philosophical thesis. As we shall see, there is no such partition: work aimed at either goal can simultaneously succeed at both. Second, the negative/positive distinction considered just by itself lumps together too many interestingly different forms of "being positive," thus obscuring the variety of ways in which x-phi work can have first-order philosophical consequences — and it is the paths to those consequences that we are looking here to taxonomize.[4]

Whereas the negative/positive distinction taxonomizes different goals of researchers, our proposal taxonomizes different paths of philosophical significance. Even philosophers who are dubious about x-phi's philosophical

---

[3] We should also point out that there are plenty of critics or sceptics of x-phi who do not follow the MM here. Even x-phi'lers themselves have raised constructive criticisms against x-phi (Horne, Livengood 2017, Rini 2015).

[4] Our work can be seen as an extension of, and complementary to, earlier work by Joshua Alexander and Jonathan M. Weinberg (2006) and Thomas Nadelhoffer and Eddy Nahmias (2007), who offer a preliminary further subdivision within the "positive" side of the distinction.

relevance generally agree that x-phi, especially in conjunction with the social and cognitive sciences more broadly, can uncover psychological facts about human behavior and cognition. On this basis, we propose a partial taxonomy of metaphilosophical bridges over the gulf between empirical results and philosophical debates. They are strategies that shepherd us from x-phi results to at least putative responses to first-order philosophical questions about knowledge, morality, human nature, truth, and the like. So, the "routes" that we consider are all ways in which psychological facts can be brought ultimately, and properly, to bear upon philosophical claims. According to such a picture, even philosophers who employ data collected by others would qualify as a kind of x-phi. Or, if one prefers, one may consider the broader category at hand to be empirical philosophy, of which x-phi is a particular, though noteworthy, sub-category. Since we are taxonomizing paths that convey us from empirical and psychological work to first-order philosophical implications, we can treat work done by philosophers and work done by psychologists, or interdisciplinary research teams, as of a kind. Philosophically relevant empirical work does not necessarily respect disciplinary boundaries, especially those boundaries that seem to be artefactual or arbitrary. What will be important for our taxonomy is that most arguments in the literature that appeal to x-phi results can be usefully situated within it, perhaps along with other empirically-based arguments.

We propose the following four routes from empirical facts, including x-phi results, to philosophically relevant consequences:

1. *Identifying* — The psychological facts are already facts of intrinsic philosophical interest. The easiest path is the path of zero distance.
2. *Tracking* — The psychological facts are evidence, albeit *prima facie* at best, for some corresponding philosophical facts. The former may be facts about what people intuit, judge, tend to say, or have implicit in their decisions, and so on.
3. *Pruning* — A psychological fact provides at least a partial defeater for a body of evidence that had been taken to support some philosophical claim. We gain knowledge by losing ignorance.
4. *Sculpting* — A psychological fact, once learned by the inquiring philosopher, in turn shapes the psychology of that philosopher's judgments, intuitions, or inferential inclinations; and thereby influences which philosophical claims will seem more or less attractive on a given set of arguments.[5]

[5] For clarity, we capitalize the names of the four routes.

In the next section, we will explain what we have in mind for each of these paths and offer a non-exhaustive set of exemplars to help illuminate them. At several points, where appropriate, we will also situate them with respect to some relevant pre-x-phi metaphilosophical work.

One last preliminary: in our role here as taxonomists, we do not mean to be either endorsing or rejecting any of the arguments or inferences as successful or unsuccessful. We are trying to offer a perspicuous manner of sorting the different x-phi animals into their cages, without offering any evaluations regarding the residents of the zoo.[6]

2. THE FOURFOLD ROUTE AND SOME ILLUSTRATIONS THEREOF

2.1. IDENTIFYING: WHEN THE PSYCHOLOGY ALREADY IS THE PHILOSOPHY

We have just emphasized the general inadequacy of the MM approach to the philosophical—empirical gulf. Nonetheless, there are at least some questions of philosophical interest that are themselves questions about human cognition and behavior. Philosophers are legitimately interested in questions about, for instance, human nature, or the consequences of believing in hard determinism, or what causal factors influence what sorts of behaviors.[7] Psychologists, of course, have an interest in the behaviors that follow from one's beliefs. These sorts of projects highlight the overlap between empirically informed philosophy and psychology.[8] Think here of the various results for trolley cases undertaken by philosophers *and* by psychologists.[9] The two may

[6] While our taxonomy could be correct even if all of the particular instances turn out to be philosophical failures, we would note that we do take it that at least a large fraction of such work does indeed succeed in achieving relevance to first-order philosophical debates. We just want to be clear that in this paper we are not looking to defend any particular piece of x-phi work.

[7] With respect to x-phi and human nature, see Knobe 2016, "Experimental Philosophy is Cognitive Science"; his very title suggests a full embrace of this zero-distance route. On the psychological consequences of beliefs about free will, see Vohs and Schooler 2008 and Nadelhoffer et al. 2020. We discuss causal factors that influence moral behavior below.

[8] See Sorrell 2017, Sytsma, Livengood 2019, and Werner 2020. Interestingly, when psychology was itself a new scholarly discipline, the prevalent angst among psychologists was its worth as a discipline of philosophic value (see, e.g., Stewart 1876). And, to an extent, what we propose in the Identifying route is a return to some of the modern empirical psychological work completed by a philosopher — e.g., Alexander Fraser (1891, 1893). For a more thorough analysis of the late modern x-phi studies, see Gonnerman, Ulatowski, Sytsma forthcoming.

[9] See contributions in Lillehammer forthcoming.

be seen as continuous insofar as empirical investigations are directly relevant to the philosophical debate.

Eric Schwitzgebel, along with his collaborators, has explored whether moral reflection positively correlates with moral behavior (for an overview: Schwitzgebel, Rust 2012). Intuitively, one would think that ethicists just are better people than those who have not had such training;[10] however, Schwitzgebel's research has shown that relatively obscure, contemporary ethics books likely borrowed by professors or advanced students are about 50% more likely to be missing from the university library than non-ethics books (Schwitzgebel 2009). Moreover, he and his colleagues have shown that ethicists do not behave any more courteously than their non-ethicist colleagues at a major philosophy conference (Schwitzgebel et al. 2012).

Not all news emanating from Schwitzgebel's lab is an indictment of ethicists behaving badly. In a monumental recent study co-authored with Bradford Cokelet and Peter Singer, using direct observational data, they found extra-laboratory behavioral effects of university-level ethics instruction. When students were exposed to a philosophy article, a 50-minute discussion section, and an optional video concerning the ill-effects of eating factory-farmed meat, students' purchases of meat of $4.99 or more at campus dining locations decreased by 7% (Schwitzgebel, Cokelet, Singer forthcoming). X-phi on the practical efficacy of different sorts of moral interventions appears to be a growth field, in fact. For example, a recent study by Buckland et al. (forthcoming) has shown how students' moral views of global poverty have been affected by engaging with Singer (1972) and others' duty argument.

2.2. TRACKING: WHEN THE PSYCHOLOGY IS EVIDENCE FOR THE PHILOSOPHICAL FACTS

At one step's remove from the cases where philosophical facts just are or simply constitute the psychological facts, we have cases where the psychological facts are nonetheless reasonably direct evidence for corresponding philosophical facts, if perhaps only ever defeasibly so. Think here of cases where a folk endorsement of P can be taken as *prima facie* evidence that P itself is true. Thus, philosophical theories that predict P gain an at least *pro tanto* advantage over theories that do not have such evidential support.

---

[10] This intuition is not universal. As an anonymous reviewer pointed out, one may claim that logicians are better people because they are more likely to follow rules. We appreciate the reviewer's comment and think that most x-phi practitioners would welcome empirical work testing their hypothesis.

The most obvious sort of case here is when there is a widely and robustly held folk intuition that P. While the easiest cases to see this route in action will generally involve explicit avowals, self-reports, or speech behaviors, in principle more implicit folk endorsement, revealed to be informing distinctions they draw or decisions they make, could also be fair game.

Here is Kripke endorsing a vague version of this idea:

> Some philosophers think that something's having intuitive content is very inconclusive evidence. I think it is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have about anything, ultimately speaking. (Kripke 1980: 42)[11]

But he does not offer any sort of proposal as to why such an evidential relation should obtain between this particular psychological fact (the intuitiveness of a content) and a putative philosophical fact (the content itself). And we doubt that philosophers should look for any one story to tell, as a matter of fact. We will sketch several here.

One way in which such an evidential relation may hold is when psychological facts at least in part *constrain* philosophical facts. Consider domains in which at least a moderate response dependence applies, such as in much of aesthetics; it is implausible, for example, that what is or is not funny, or beautiful, or a horror flick can outstrip what people think and say and feel about those categories. One should similarly expect theses in the philosophy of language to be partially bounded, even if not directly determined, by linguistic behavior; or in political philosophy, a theory of justice may be informed, though not dictated, by just how cold-bloodedly self-serving humans really are. In such cases, finding that people's assertions or behaviors most closely comport with P, provide evidence that the correct philosophical theory will be one that is consistent with P.

We might also take ourselves to have reason to think that human speech, cognition, or action is evidence about corresponding philosophical facts not because they are metaphysically connected to and constraining those facts but because we expect the former to be reliably tracking the latter. Such a view is promoted by J. L. Austin in his famous apologia for ordinary language philosophy:[12]

---

[11] See Deutsch 2015 for what kinds of intuitiveness Kripke has in mind.

[12] Some experimental philosophers take themselves to be flying under the Austinian flag — e.g., Fischer 2014, Fischer et al. 2019, Hansen and Chemla 2015, Hansen, Porter, Francis 2019, Buckwalter 2010, and Knobe 2003. Interestingly, Austin did not have a positive opinion of philosophical projects that employed experimental studies. The closest ally here is Næss' empirical semantics. Austin believed that his ordinary language philosophy was "superior" to Arne Næss' empirical semantics. We learn from Geoffrey Warnock that Austin

> our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our armchairs of an afternoon — the most favoured alternative method. (Austin 1979: 182)

He goes on a couple of pages later:

> [Ordinary linguistic practice] has been derived only from the sources available to ordinary men throughout most of civilized history: it has not been fed from the resources of the microscope and its successors. . . . Certainly, then, ordinary language is not the last word: in principle it can everywhere be supplemented and improved upon and superseded. Only remember, it *is* the *first* word. (Austin 1979: 185)[13]

It is too bad, if completely understandable, that he did not consider the possibility of a more experimentally ambitious set of methods! Even some philosophers who do not wholeheartedly endorse the Austinian program in all its particulars have nonetheless embraced something like this story as a justification for their armchair-bound activities.[14]

"was careful to distinguish the programme he had in mind from the kind of Gallup-poll, empirical teamwork which Næss believed in, and which Austin regarded as, in principle, misguided" (1969: 14, n. 2, Urmson, Quine, Hampshire 1969: 80). Austin seems to endorse the MM metaphilosophical position that we mentioned earlier. For a more updated Austinian critique of experimental philosophy, see Baz 2018.

[13] Austin is of course not the first to reason along these lines. E.g., A. T. Shearman in the early 20th century wrote that philosophers drawing upon psychology "would give wider and more enduring satisfaction to the human mind . . . [because] the intuitions of the majority would be the starting point instead of the intuitions peculiar to isolated thinkers. . . . It will be observed that I do not expect finality of philosophical doctrine to be reached in this way. I think there is no possible finality. I merely submit that the system evolved on this plan would be likely to produce more good than would be produced by several systems antagonistic to one another." (Shearman 1907: 160-161)

[14] See the very title of Keith DeRose's 2005 paper, "The Ordinary Language Basis for Contextualism, and the New Invariantism," echoed in the x-phi follow up by Wesley Buckwalter (2010), "Knowledge Isn't Closed on Saturday: A Study in Ordinary Language." The "Canberra plan" perhaps comes by its embrace of the folk more by way of Lewis than by Austin, and it is not unreasonable to think that *a* possible use of x-phi is to provide a naturalistically respectable method for pursuit of that program. For an extended discussion of this, see Ulatowski 2017: 63-68, Ulatowski ms. Some slightly later philosophers appeal not so much to common language but common concepts (e.g., Goldman 1979).

We would also suggest that, in addition to such articulated metaphilosophical appeals to common folk intuitions, one can plausibly locate it as implicit in much if not most analytical philosophy since the middle of last century (as illustrated, e.g., by Black 1950). See Weinberg 2016a on how generally philosophers have not been explicit regarding the metaphilosophical grounding for their intuitive evidential practices.

We can see similar metaphilosophical thinking emerging in other streams of the analytic tradition. Working around the same time as Austin was Rudolf Carnap (1955) on his notion of intensionality in natural languages. After providing a brief summary of an empirical procedure for determining intensions, Carnap admits that one could derive philosophically interesting facts about intension by "testing hypotheses concerning intensions" (1955: 40). "Therefore the rules could be understood and carried out by any scientist (provided he is not infected by philosophical prejudices)" (Carnap 1955: 40). In a footnote, Carnap cites the work of Næss who has performed empirical studies in a way that Carnap believes they should be carried out. He writes:

> [Arne Næss' *Interpretation and Preciseness*] describes in detail various procedures for testing hypotheses concerning the synonymity of expressions with the help of questionnaires, and gives examples of statistical results found with these questionnaires. The practical difficulties and sources of possible errors are carefully investigated. The procedures concern the responses of the test persons, not to observed objects as in the present paper, but to pairs of sentences within specified contexts. Therefore the questions are formulated in the metalanguage, e.g., "Do the two given sentences in the given context express the same assertion to you?" Although there may be different opinions concerning some features of the various procedures, it seems to me that the book marks an important progress in the methodology of empirical meaning analysis for natural languages. . . . The book, both in its methodological discussions and in its reports on experiences with the questionnaires, seems to me to provide abundant evidence in support of the intensionalist thesis. (Carnap 1955: 46, n. 6)

In its current manifestation, the Tracking route may be traced to a great deal of the experimental work completed since 2001. Many classic experimental philosophy papers can be seen as exemplifying this path — for example, the experimental works on intentional action (Adams, Steadman 2004a, b, Cushman 2007, Knobe 2003, 2004, Machery 2008, Nichols, Ulatowski 2007, Paprzycka-Hausman 2015), on free will and determinism (Nahmias et al. 2006, Nichols, Knobe 2007, Nichols, Roskies 2008), on epistemic intuitions (Feltz, Zarpentine 2010, May et al. 2010, Sripada, Stanley 2012), or on semantic intuitions (Næss 1938a, b, 1953, Machery et al. 2004).

We hasten to add that this second route is in no way restricted just to inferences from *folk* cognition to philosophical facts. One may instead look to make inferences from the behavior or cognition of a cognitively elite subset of people, to a relevant philosophical truth. There are a range of ways in which a group might plausibly be elite, in the sense that their psychology on the whole may better reflect the relevant philosophical reality than the psychology of humans in general; training (such as a specialist education training in the topic at hand) or cognitive traits (such as being more reflective and thoughtful) are good candidates. There are indeed some cases where non-philosophers

are simply ineligible for the task — there is little point in asking non-specialists about, say, different physically or functionally reductive accounts of what a gene is,[15] or about the a priori / a posteriori distinction in those terms. Outside of such truly and legitimately highfalutin tasks, though, we do think experimentalists would be better off pursuing parallel researches of both folk and putative elites, since the jury is still very much out as to just which kinds of characteristics provide what kind of improvements on which kind of tasks.[16]

2.3. PRUNING: WHEN THE PSYCHOLOGY UNDERMINES SPECIFIC PHILOSOPHICAL EVIDENCE

Yet a further step can be taken, while maintaining a bridge from x-phi to first-order philosophy. Here we have cases where experimental findings do not themselves serve as direct evidence for a philosophical thesis, but as indirect evidence, by providing an empirical defeater for some piece of countervailing evidence. When misleading evidence is allowed to persist or even proliferate, it can obstruct inquiry by reducing our credence in true hypotheses — hypotheses whose truth might have been more apparent based on other lines of reasoning, were it not occluded by the misleading evidence. Undermining problematic pieces of evidence can thus constitute a substantive contribution to ongoing debates and offers the promise of moving debates forward that may otherwise seem irredeemably stuck (see Sytsma 2010, Sytsma, Livengood 2012).

This route from psychology to philosophy is clearly related to the negative program, as can be seen in its resonance with the construal of the negative-program philosopher as one "who advocates not the root and branch removal of all intuitions, but just the pruning away of some of the more poisoned philosophical branches" (Alexander, Weinberg 2006: 71). But what is advocated in that earlier paper is a fairly sweeping culling that would remove, at a minimum, all the "peculiar and esoteric intuitions" (Alexander, Weinberg 2006: 71) *en*

---

[15] See Stotz, Griffiths, Knight 2004 for a fine example.

[16] See, e.g., the contrasting pictures that N. Ángel Pinillos et al. (2011), Weinberg et al. (2012), and Mikkel Gerken and James Beebe (2016) paint on the impact of reflection on philosophical case verdict tasks; and see Kneer et al. (forthcoming) for a more forceful empirical statement. The literature on training and expertise in philosophy is both increasingly large yet highly inconclusive, because it has largely been more theoretical than straightforwardly experimental; see Nado 2014, 2015 and Alexander 2016 for sceptical reviews, with Irikefe 2020 for some pushback. We would also note that it is not just cognitive variation among people but also across different circumstances, settings, or practices that might yield a more appropriate basis for philosophical inference; for reasons of space, we do not pursue those here.

masse. In this paper, we have in mind a much more targeted and careful cropping: it is more like sculpting a bonsai tree than defoliating an entire forest.

Another difference between the Pruning route and the negative program can be seen when we consider what the philosophical upshots of each are meant to be. Consider a debate between theories A and B that has been bogged down because (i) general considerations speak in favor of A, but (ii) B does a much better job handling a particular class of recalcitrant intuitions. Now suppose that some x-phi studies give us reason to distrust those B-supporting intuitions *in particular* — that is, not as part of a more general critique of intuitions. This targeted specificity of Pruning is what makes it not just a restatement of the negative program, which aims instead to undermine armchair intuitional practices more generally.[17] Furthermore, the negative program tends not to offer positive philosophical upshots — that is why it is the negative program, after all. But as a consequence of this hypothetical instance of Pruning, we may place ourselves in a good position to say that the x-phi tilts the overall body of evidence into A's favor, even though the psychological results are not themselves any sort of direct evidence for A. They undermine the support for B, and thus indirectly give us reason to endorse B's rival.

A clear precursor of, and inspiration to, the x-phi-based arguments that would emerge just a couple of years later can be found in Tamara Horowitz's paper (1998). She considers intuitions, widely reported, that although in many situations it would be morally required to save five people at the cost of not saving one, in otherwise similar situations it is forbidden to kill one in order to save five. Such intuitions are problems for the consequentialist, who will want to score both such scenarios in similar terms (5 – 1 = 4). She suggests that psychological work on loss aversion may lead us to discount such intuitions as artifacts of a biased cognitive system (cf. Kahneman, Tversky 1982). Her hypothesis is that when someone needs saving, they count as in some sense already gone, and so simply not saving them is not a further loss; it is just an unrealized gain. But to kill someone would mean a loss of a life that was not already going to be lost, and thus it is a target for loss aversion. The psychological facts about loss aversion are not themselves any sort of evidence for consequentialism. But they undermine an important piece of evidence against it, and thus support consequentialism indirectly.[18] We are not endors-

[17] See Nado 2017 on the value of targeted intuitional challenges; see also Weinberg 2016b on "going positive by going negative."

[18] Jonathan Ichikawa (2009) provides an elegant discussion of this and several other examples of "explaining away." He also raises interesting questions about the applicability of classic negative program cognitive diversity findings, but since, as we have emphasized, the negative program is not our particular concern here, we will not pursue it.

ing Horowitz's argument; rather, we are using her work as an example of the kind of approach that we have in mind. Her argument thus exemplifies our third route from psychological facts to philosophical results in action.

Another nice illustration comes from recent work by Felipe DeBrigard on Robert Nozick's "experience machine" thought-experiment (2010; cf. Weijers 2016). In Nozick's original argument (1974: ch. 3, 1989: 104), we contemplate whether we should choose to embed ourselves, permanently, into a virtual world that was as perfectly pleasurable to experience as we could imagine — except, it is all just a programmed hallucination, including as part of the hallucination that we utterly forget that it is a computer-generated simulation. We have experiences of falling in love or triumphing over adversity, but none of it is truly happening. Nozick suggests, and philosophers have widely agreed, that we would not choose to enter irrevocably into the experience machine (cf. Hewitt 2010, Silverstein 2000). Since we would choose a less-pleasant reality over a more-pleasant hallucination, we seem to be naturally choosing in a way that implicitly rejects experiential hedonism. DeBrigard, motivated by the widely confirmed phenomenon of status quo bias, turned Nozick's vignette around and asked the following of study participants: if you learned that you have been in an experience machine all along, would you choose to be extracted from the (virtual) life as you have known it, and placed into reality? If people have an inclination to maintain the status quo, then they would choose to remain in the experience machine since that is the life they have been a part of all along. DeBrigard's experimental participants, confronted with that hypothetical, strongly preferred to stay in their accustomed virtual unreality.[19] This overall pattern — choosing to stay in reality when starting in reality, but choosing to stay in a simulation when starting in a simulation — suggests that what drives the intuitions in these cases is less something deep about hedonism, and more something shallow about anchoring (status quo bias) and loss aversion (Knetsch 1989, Samuelson,

[19] DeBrigard's experiments included a positive, neutral, and negative scenario. In the positive scenario, the respondents are told that in real life they are "multimillionaire artists living in Monaco." In the negative scenario, respondents are told that in real life they are incarcerated in a maximum-security prison. In the neutral scenario, respondents are told they would return to their own real life. Study participants were assigned to one of the scenarios and asked whether they would "Remain connected" or "Go back to reality." In response to the negative scenario, a great majority of respondents (87%) chose to remain, while in both the neutral and positive scenario participants were divided relatively evenly between going back (Neutral: 54%; Positive: 50%) and remaining (Neutral: 46%; Positive: 50%). DeBrigard believed that these results showed that contact with reality was not nearly as critical as Nozick's original thought experiment had suggested.

Zeckhauser 1988). The philosophical payoff is primarily subtractive, in proposing a defeater for Nozick's original argument, but no less real for that.[20]

We will close this section with one more prominent example of how Pruning has been a part of experimental work in philosophy and psychology. Joshua Greene (2008, 2013) has demonstrated the existence of both apparently consequentialist and deontological ethical intuitions in study participants' evaluations of trolley cases and the like. That result, in itself, would be unsurprising to philosophers. Greene's "pruning" style argument is rooted in his further psychological claim that the putatively "Kantian" elements in our intuitive moral psychology are not based on reason but on emotion, as revealed by comparing the brain scans of participants making judgments about trolley-type cases under more or less emotionally engaged conditions, and also comparing the scans of those who make the consequentialist choice and those who refrained from doing so.[21] According to Greene, the affective basis of those intuitions gives us reason to doubt their evidentiary value — thus, pruning away the recalcitrant anti-consequentialist intuitions and shifting more evidentiary weight behind consequentialism. We will not rehearse the details of that part of his argument, which has of course proved controversial; our purpose is not to endorse the soundness of any of the arguments that we discuss but only to locate their manner of philosophical inference in our proposed taxonomy.

2.4. SCULPTING: WHEN THE PSYCHOLOGY CHANGES THE PHILOSOPHER'S MIND

Much of the potential in the Tracking and Pruning routes presuppose that the intuitive verdicts about philosophical vignettes can contribute some sort of evidence for philosophical inferences. As we saw, some Tracking projects are meant to address the question of what the distribution is of various intuitions in the population at large, such as the attribution of free will in a deterministic universe. Some Pruning projects aim to offer re-descriptions of the crucial factors driving our intuitive responses in some cases, such as in Nozick's experience machine case, in a way that would shift what conclusions we might take them to be evidence for. In addition to changing our understanding of what intuitions are out there or what causes them, x-phi may also

[20] This sort of "progress through undermining" further illustrates the positive element in Pruning that we take to distinguish it from the negative program. Of course, empirical defeaters can themselves be undermined, with contrary experimental findings, see, e.g., Rowland 2017. Our botanical metaphor probably breaks down at this point, though, since one cannot restore a pruned limb by "counter-pruning"!

[21] See also Kelly 2011, 2018, on the untrustworthiness of disgust-based moral intuitions.

be able to change what intuitions we end up having in the first place. (Or, rather, in the second place.)

The idea that our judgments about cases can be non-inferentially changed goes back at least to Nelson Goodman's *Fact, Fiction, and Forecast* (1983), though arguably there is consideration only of other cases and principles within a normative domain that apply the pressure for alteration (cf. Ulatowski 2017: § 2.3).[22] More recent philosophers have noted that our judgments about cases may be driven by our background empirical picture of the world, and hence also be modified as that picture is updated.

For example, Ned Block and Robert Stalnaker point out that even putatively a priori philosophical judgments about a hypothetical case may be shaped by the reasoner's understanding of the available scientific evidence. They write:

> we should reflect on what we should (not would but would if rational) say if we were to find out certain things about the actual world. If we were to find out that the colorless odorless drinkable (more or less) stuff in rivers and lakes is XYZ, we would conclude that water is (necessarily) XYZ. If we found that such stuff is not really a liquid, we would have found out, and we should say, that water is not a liquid. (But isn't it also true that if we were to learn that the word 'coumarone' referred to an extinct flightless bird, we would conclude that coumarones — as we would put it — are extinct flightless birds?) This seems to be armchair reasoning, reflection that does not include any obvious reference to real experiments, so it is tempting to conclude that this reflection just unfolds our concepts in a totally a priori way. But what this conclusion misses is that our reasoning about the proper epistemic response in various counterfactual situations is informed not only by our concepts, but by implicit and explicit theories and general methodological principles that we have absorbed through our scientific culture — everything that the "we" who are performing these thought experiments believe. What people should rationally say in response to various hypothesized discoveries will vary depending on their experience, commitments and epistemic priorities. (Block, Stalnaker 1999: 42-43)

Block and Stalnaker's suggestion is well taken. Even when we are engaged in what we take to be an a priori argument, we may nonetheless be employing empirical evidence that has been "absorbed" by us, whether via the culture at large or by means of a more individualized course of reading or observation.[23]

---

[22] The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences, and in the agreement achieved lies the only justification needed for either. The method of reflective equilibrium, thus, is a matter of bringing into accord judgments about particular inferences and about general principles of general inference. The accord gained is the only justification possible for the inferential principles that emerged. John Rawls, interestingly, extended such a method to moral principles and moral judgments (1971: 20). See Daniels 1979 on wide and narrow reflective equilibrium.

[23] See also Williamson 2005: 14: "Very often, the background knowledge needed to evaluate a counterfactual consists not of specific items of information acquired on specific occasions but of a more general sense of how things go, honed over long experience."

Taking things meta for a moment, we would note that this picture of human judgment's recruiting of empirical background knowledge is itself shaped by learning about scientific psychology. Thus, to the extent that internalizing that picture in turn implicitly shapes one's judgments about such matters as the epistemic scope of philosophical argumentation — say, the attractiveness of a theory of intuition as a kind of pure rational apperception — we have a further illustration of this route from psychological facts to philosophical results.

While Block and Stalnaker emphasize theory and general methodological experience, specific empirical findings, too, may non-evidentially mold our philosophical judgments. More than a decade before "Normativity and Epistemic Intuitions" (Weinberg, Nichols, Stich 2001), Stephen Stich (1988) had floated the hypothetical possibility of different social groups having different epistemic intuitions and deployed that possibility to try to raise concerns about the method of reflective equilibrium, since in principle, different starting intuitions should lead to different equilibria. Yet his then-hypothetical arguments evoked only theoretical or equally hypothetical responses from the literature. The putative cognitive diversity findings of (Weinberg, Nichols, Stich 2001) and (Machery et al. 2004), however, raised a rather greater ruckus, even though those studies, as their authors stressed, were small and very preliminary (cf. Kim, Yuan 2015, Machery et al. 2017). It simply cannot be that the philosophical community's priors regarding substantial cognitive diversity were rationally raised by these findings. Rather than treating the philosophical community as collectively irrational in this regard, we prefer to think that those results from x-phi's infancy evoked a subterranean realignment of their judgments about how pressing various hitherto-ignorable concerns might be.

A second example here might be the influence that experimental philosophy has upon what has come to be known as "conceptual engineering." There is already quite a bit of work that connects the potential descriptive import of x-phi data to Carnapian explication (see Koch 2019, Lindauer 2020, Nado 2021, Pinder 2017a, b, Schubach 2017, Shepherd, Justus 2015). However, we contend that x-phi may have normatively important connections with conceptual engineering — which is, after all, centrally concerned with crafting new concepts for our use. X-phi may not only provide some descriptive basis for reshaping our thinking, but in addition it may provide some measurement and prescriptive guidance for which options to choose on the basis of experimental data. Suppose that we have proposed a new concept of knowledge under consideration, call it K*. We would want to determine the costs and benefits of our thinking in terms of K*, instead of our current concept K; we might need a better sense of just how different K* really is from K; how easily

K* could be promulgated and adopted; whether K*'s adoption would have the positive effects on inquiry or politics that were desired; and so on. Such questions about how to shape our concepts may thus require a substantial contribution from x-phi, perhaps among many relevant descriptive sources, throughout the successive phases of a conceptual engineering project, roughly: description, evaluation, improvement, implementation.[24]

A third example may be drawn from a possible empirical challenge to the primitivist theory of truth according to which "the concept of truth is fundamental: it is explanatorily indispensable concept that cannot be defined, analyzed, or otherwise explicated in terms of concepts that are still more fundamental" (Asay 2021: 525). Perhaps the most effective argument for conceptual primitivism is the omnipresence argument, which depends on the concept of truth contributing to the structure of every propositional thought. Asay gives us a clue as to where we might look for a challenge to the omnipresence argument. In a telling footnote, Asay writes:

> my view has empirical implications, in that it's an empirical matter as to which creatures possess which concepts. It's not straightforward, however, that omnipresence can be directly tested by empirical methods. The falsifier would be someone who could assert, believe, and contemplate but didn't possess TRUTH. But there is no independent means of identifying who possesses a certain concept in the absence of (philosophical) views as to what it is to possess that concept. (Asay 2021: 537, n. 8)

There is no doubt that asserting, believing, and contemplating are signs of advanced cognition. Additionally, understanding the behavior of others in a wide variety of situations requires an understanding of their psychological states. So, a critical marker of advanced social cognition would be an understanding of false beliefs. Nonverbal false-belief tests using active behavioral measures have shown that — by differing in their helping behaviors — infants between 16- and 18-months old demonstrated their ability to attribute different goals to an experimenter based on his true versus false beliefs (Buttelmann, Carpenter, Tomasello 2009). False-belief attribution by human infants suggests that, while perhaps nonverbal, they display behavior consistent with asserting, believing, or contemplating what others believe or fail to believe. By Asay's own lights, then, it is a short step from here to the possession of TRUTH.

For infants it seems reasonable to suppose that they assert, believe, or contemplate some propositional content *and* possess TRUTH, but we may not be compelled to allow such an inference to go through for non-human animals, like apes. In a recent study, David Buttelmann, Frances Buttelmann,

---

[24] See Andow 2020 and, especially, Koch 2020 on the principle of "divergence."

Malinda Carpenter, Josep Call, and Michael Tomasello (2017) do exactly that. They show — using the new interactive behavioral infant false-belief test pioneered in (Buttelmann, Carpenter, Tomasello 2009) — that apes may have a basic understanding of others' false beliefs. The pattern of results was exactly the same as that found in the study on 16-month-old infants.

While one may enthusiastically support the idea that an infant possesses TRUTH, one may be more reluctant to ascribe the possession of TRUTH to a chimpanzee.[25] Even philosophers who are willing to accept that philosophy should be empirically informed may be reluctant to accept such ascriptions because they would not only have to ascribe TRUTH to apes but might also have to accept that non-human animals are in possession of it and other primitive concepts, like IDENTITY, EXISTENCE, KNOWLEDGE, GOODNESS, BELIEF, ACTION, and CAUSE.[26]

3. "BUT SURELY NOT…": X-PHI AND PHILOSOPHICAL LOGIC

Although we hope that the preceding sections have made clear that x-phi may be fruitfully understood as contributing resources to a wide range of philosophical questions in at least the four ways that we have discussed, we also want to make a case for the applicability of x-phi to philosophical projects in areas that do not immediately seem to be amenable to such contributions. In doing so, we hope to illustrate the four paths themselves and to underscore just how far — and how deep — philosophy may be able to travel along them.

---

[25] If one is worried that this is not experimental philosophy in the strictest sense, we refer the reader to section 1, where we indicated that, for our purposes here, it is not important to distinguish x-phi from empirical philosophy more generally.

[26] Ascribing primitive concepts to non-human animals might make some a tad skittish, but notice how the philosophical argument here has bottomed out at a question that is neither a priori nor quite amenable to being well settled by experimental methods. We cannot rule out that non-human animals possess primitive concepts by *fiat* just as we cannot wholeheartedly accept that they do possess such concepts through experimentation. This is where the fourth route becomes salient: coming to learn these findings about sophisticated capacities of non-human animals, and many, many more results like them (see Horschler, MacLean, Santos 2020 for a recent review, including discussion of the experimental complications in this area) can have a profound effect on a philosopher's inclinations to accept or withhold an attribution of such concepts to other primates. The skittish feeling cannot be removed evidentially via counterargument. But it *can* be treated, as the cognitive uptake of scientific findings work their way into our broader judgments as indicated by philosophers like Block, Stalnaker, and Williamson.

What kinds of questions can x-phi help us to resolve? Many philosophical questions seem to be too esoteric to be useful for the life we live. This is why the first-order psychological and behavioral facts speak to questions of, for instance, moral reasoning and even aesthetics. Likewise, empirically informed characterizations of human faculties and capacities are *prima facie* relevant to questions about humans as knowers and thinkers. But what about such seemingly esoteric questions like: what is the nature of truth? While the average person may want to know whether a statement is true or accurate, they do not necessarily have a concern for what truth is. Or what is the proper resolution of semantic paradoxes? These are *loci* where our cognition has apparently tied itself in a knot — so other than learning that, indeed, the human mind is capable of auto-entanglement, what is to be learned from studying such self-gnarled minds? These seem to be areas where human psychology might be part of the problem, where to succeed in our endeavors we must transcend our minds and not collapse back inside them. Let us turn to three problems from philosophical logic that seem beyond the pale: the nature of truth, the Liar Paradox, and vagueness. In each case we will suggest that x-phi, understood in terms of the four routes, might support our efforts and help to distinguish between viable and non-viable approaches.[27]

3.1. THE NATURE OF TRUTH

Empirical psychological results on the nature of truth pre-date x-phi as we have come to know it over the last decade or so. In the mid-1930s, Arne Næss performed a series of quantitative and qualitative experimental studies on the non-philosopher's concept of truth (cf. 1938a, b).[28] Næss used open-ended questions, for instance:

> What is to be understood by the expression "something is true"? Define the expression. (Næss 1938a: 24)
>
> What is the c.c. (common characteristic) of that which is wrong? (Næss 1938a: 23)
>
> Give me an example of something that is true. (Næss 1938a: 23)

---

[27] We would like to note that Jonah Schubach has a wonderful experimental research program on questions that seem to be equally esoteric. See, e.g., Schubach 2017, 2018.

[28] Notably, work that comprised Næss' middle period, circa 1950-1972, was far more sophisticated in its experimental design and statistical analyses. We would not like to get bogged down in the details here; rather, we urge the reader to return to Næss' work and to some recent studies in the secondary literature (Chapman 2011, 2018, Ulatowski 2016, 2017, 2018).

> Do you employ the expression "the truth"? (If answered positively:) On which occasions? (Næss 1938a: 26)

An assistant recorded each subject's response to these questions, and the data were analyzed. The questionnaire method, though fruitful, did not lead Næss to a uniform view of the non-philosopher's notion of truth. Næss writes:

> We have gathered more than 1000 examples from non-philosophers and a great many from philosophic literature, but it is by no means plain how we from this collection should be able to infer any *general* statement resembling definitions. (Næss 1938a: 71)

Given the wide variety of ordinary notions of truth he collected using the questionnaire method, none of them stand alone as *the* common-sense view. Even if there is not a singular and uniform notion of truth compatible with all of the results collected by Næss, his data are interesting at least because he has begun to get a better sense of the intuitive terrain. Næss approached the project ecologically, permitting us to better appreciate and understand the topographical contours of non-philosophers' intuitions about the concept of truth (cf. Ulatowski 2017: 63-68).

While responses from the folk do not settle philosophical questions about the nature of truth, they might help to support or undermine one or another philosophical proposal. But the less appreciated importance of an ecological approach can be seen from the effect that Næss' work had upon the development of others, most notably Tarski. Barnard and Ulatowski (2016) argue that the development of Tarski's thinking about the adequacy of the formalized definition of truth was influenced by his awareness of Næss' results. Barnard and Ulatowski take notice of how Tarski's papers on truth, logical consequence, semantic concepts, and definability identify two primary conditions for successful definitions: formal correctness and material (or intuitive) adequacy. Material adequacy requires that the concept expressed by the formal definition captures the intuitive content of truth. Tarski's thinking about material adequacy and the content of the common-sense, every-day usage of truth developed over time from 1933 to 1944, between his two most important contributions on the concept of truth (Tarski 1933/1983, 1944). In the "Polemical Remarks" of Tarski (1944), especially §§ 14 and 17, Tarski's references to Næss' empirical research show that Tarski's thinking about the everyday conception of truth went from complete rejection of its relevance to a definition of truth to a respect for Næss' work, especially his findings with respect to Tarski's definition of truth and his research that suggests there is no single common or everyday concept of truth (Tarski 1944: 374, n. 29). In effect, Tarski's thinking about truth was updated to reflect the new empirical data. This is just one example of the fourth route in action.

So, where and how does Næss' ecological approach fit into the fourfold route? His work does not align philosophy with psychology in the way that we have outlined in the Identifying route. Asking study participants qualitative questions about their view of truth is not also a question about human cognition and behavior. Similarly, Næss is not necessarily showing how others' empirical work has failed to appreciate what the non-philosopher's view of truth is, even though in the opening pages of his (1938a) there is an indictment of philosophers who have failed to even consider what non-philosophers think about truth. So, we can say that his view is not a reflection of either the Pruning or Sculpting route. He certainly was not able to convince contemporaries of the relevance of his empirical work for philosophy (e.g., Moore 1939, Nagel 1939). By process of elimination, it should then be that Næss' work is an example of the Tracking route. This is at least partly true. Næss found direct empirical evidence for corresponding philosophical theories of truth, such as the correspondence theory and disquotationalism. However, he did not find that there was one correct theory of truth that was supported by the population of study participants generally. What then to do with the data became a challenge for Næss and for the Oslo school of empirical semantics.

3.2. THE LIAR PARADOX

The Liar Paradox is an all too familiar problem with hotly contested possible solutions. Some sentence or set of sentences directly or indirectly asserts of itself that it is false. If the sentence that says of itself that it is false is true, then the sentence is strictly speaking false; but, if the sentence is false, then if it says of itself that it is false and it is false, then it is true. If *L* is the liar sentence: "*L* is false," then we might ask whether *L* is true or false. Perfectly ordinary reasoning leads us to recognize that *L* may not be *sensu stricto* true or false. So, we might be allowed to imagine that "*L* is neither true nor false" (gappy) or "*L* is both true and false" (glutty). If we treat *L* as being true or being false, or as being "glutty," the result is a contradiction. If we opt for the "gappy" option, then we appear to concede another logical principle: the Law of Excluded Middle.

It is safe to say that the liar sentence, *L*, tends to arouse a sense of puzzlement in scholars and students alike, but it is clearly an empirical question how people actually process such paradoxical cases. This is where it makes sense to turn to x-phi. As David Ripley pithily noted:

> We should expect experiment and logic to fruitfully interact whenever a field of inquiry involves rigging up a logical system to capture some experimentally-explorable phenomenon: in these cases, logical approaches will help us decide which aspects of the

> phenomenon to experimentally explore, and experimental approaches will help us choose which logics best capture the phenomenon. (Ripley 2016: 532)

The aim of our earlier paper (Barnard, Ulatowski, Weinberg 2017) on the Liar Paradox was to employ experimental methods to assess how and why the folk understand self-referential sentences such as *L* to fall into one or the other of the four cases noted above.

In that study, we asked participants to categorize the demonstrative form of *L*: "This sentence is false" into one of **true**; **false**; **neither true nor false**; or **both true and false**.[29] When these responses were analyzed, a clear majority of respondents chose **neither true nor false**, both to the demonstrative form of *L* as well as to an alternate version: "The Sentence in the box is false." The **neither** response was more likely to be found among philosophically sophisticated respondents, those who indicated a commitment to the law of noncontradiction, and those whose responses indicated higher levels of reflective thinking. In particular, high performance on either the Wason Selection Task or the Cognitive Reflection Test correlated with an increased likelihood of answering **neither**. Vanishingly few subjects chose **true**. We conjectured that the **false** response was most likely a product of participants' heuristic reasoning, since the sentence does say of itself that it is false.

We can see at least three of the four routes to philosophical relevance available from these results. First, it seems to us a good candidate for being of intrinsic philosophical interest just how the folk do try to handle paradoxes when confronted with them (the Identifying route). In that paper, we identified some philosophers positing that the folk notion of truth is simply incoherent, in and of itself, for example:

> The significance of a paradox is never the paradox itself, but what it is a symptom of. For a paradox demonstrates that our understanding of some basic concept or cluster of concepts is crucially flawed, that the concepts break down in limiting cases. And although the limiting cases may strike us as odd or unlikely, or even amusing, the flaw itself is a feature of the concepts, not the limiting cases that bring it to the fore. If the concepts are important ones, this is no laughing matter. . . . if our concept of truth is somehow incoherent, as the paradox suggests, this raises the question of whether the same incoherence infects the mathematical and scientific discourse that presupposes the intuitive notion. (Barwise, Etchemendy 1987: 4-5)

This seems to predict that folk responses may simply fall apart in the face of the paradox, and this contrasts with other philosophers (Maudlin 2008) who postulate that non-specialists will have coherent ways of responding, and the challenge to the specialist is to work out the formal details of how such re-

[29] Boldface is introduced here to indicate when we are using the term to name one of the responses on the instrument.

sponses can be systematized and make larger theoretical sense. Our 2017 paper can thus be situated on the "folk coherence" side of that debate.

Our 2017 arguments also exemplify the Tracking route, where the pattern of findings gives *prima facie* evidence towards one side of a philosophical debate. In general terms, these results suggest that in paradoxical cases the folk and philosophical experts alike are contradiction averse to the extent that they are willing to set aside the presumption in favor of bivalence. Moreover, the participants who are more likely to be reliable judges on matters of logic and truth — the more reflective and the more philosophically trained — showed a greater tilt in that direction.[30] This is not, of course, definitive of how to properly understand the Liar Paradox, but it suggests that the ordinary appreciation of the paradox is not of a case where a sentence is supposed to be understood as both true and false, but rather as neither. Therefore, attempts to offer plausible philosophical accounts of the paradox should probably try to fit this model where truth value gaps are preferred to truth value gluts.

The fourth, Sculpting route is where philosophers' intuitions or reasoning patterns in a target philosophical domain are themselves shaped by learning about experimental results. One can easily find armchair versions of the Sculpting strategy regarding the possibility of true contradictions, or at least, the imaginability of them — namely, in the use of colorful and vivid philosophical fictions, such as Graham Priest's (1997) "Sylvan's Box" or Tamar Gendler's (2000) "Tower of Goldbach."[31] By engaging with such fictions, the philosophers who find them at least apparently conceivable may also find their intuitions regarding true contradictions somewhat "softened up."

We would suggest that fiction is not the only genre that can work on philosophers' intuitions that way, but that experimental reports can potentially do so as well — including, perhaps, in the opposite direction of the fictions. One of us reported that he had long found the **both** reply, and with it the rejection of the law of noncontradiction, not *intuitive*, but *attractive*, especially upon reading the stories cited above. But the fairly robust folk rejection of such responses to *L* has made dialetheist approaches lose some of their shine.[32] Antecedently, he would not have thought that his experience of the attrac-

---

[30] See our discussion of "cognitively elite subsets" above (section 2.2).

[31] See also Meskin, Weinberg 2012 for a few further such *prima facie* impossible-but-imaginable short stories, as well as some speculation regarding the psychological processes involved in our engagement with such stories.

[32] However, it would be very interesting to run a study to put such stories before a pool of subjects and see if they are more willing to embrace contradictions with all that narrative scaffolding.

tiveness of different options would have depended in any way on the folk's responses. Yet the experience of learning the results, perhaps cumulative upon the experience of participating in designing and running the study, apparently impacted his cognition along the lines of the fourth route. It is not that he took the folk responses as evidence against dialetheism; rather, they seemed to have prompted an episode of self-sceptical reconsideration. We do not wish to make too much of this anecdote, which is after all a lone biographical episode. Our point is just to observe it as a possible instance of the fourth route in action, so as to illustrate what such a route might look like in the context of philosophical logic.

3.3. VAGUENESS AND THE SORITES PARADOX

Vagueness is another phenomenon that has been subject to philosophical, psychological, and logical analysis. At some level, it is similar to the Liar Paradox; the ordinary application of vague categorization generates its own family of "sorites" paradoxes. Vagueness as a general phenomenon is encountered in a wide range of situations — for example, the demarcation of physical boundaries (as with a coastline, the edge of a cloud, or where a valley ends and a mountain begins) and familiar ("tall," "bald," etc.) and abstract ("is a small number") cases of classification. Vagueness as a target of philosophical interest, since at least Bertrand Russell (1923), has been understood to turn on the problem of "borderline cases," which are situations where the application of a vague term or vague predicate and its opposite or complement both seem to be permissible given the relevant norms governing the use of the term or predicate. "Tall" is a commonly noted vague term in this sense. So, if a given man is neither determinately tall nor determinately short but falls in between, then such a man may be a borderline case of tall. So, it appears permissible to say of the man that he is tall and that he is not-tall.

A number of recent studies have tried to assess how folk respond to borderline cases. Nicolao Bonini, Daniel Osherson, Riccardo Viale, and Timothy Williamson (1999) performed a series of experiments that aimed to characterize the "unstable" character of vague terms by asking respondents to complete prompts such as:

> It is true [false] to say that a man is tall [not tall] if his height is greater than [less than] or equal to ___ cm.

The pattern of responses showed that the mean highest value for, say, "not tall" and the mean lowest value for "tall" were significantly different and non-overlapping. The study, thus, confirmed their hypothesis that the vagueness or

unstable character of these vague concepts resulted from a kind of ignorance about where the real boundary between them is, which is consistent with an "epistemicist" position defended by one of the authors (Williamson 1994).

Following on from the study of Bonini et al. (1999), Sam Alxatib and Francis J. Pelletier (2011) performed a study where participants were shown an image of a police line-up with five numbered male figures of varying heights. For each figure, participants were asked whether the following four claims: "#*x* is tall," "#*x* is not tall," "#*x* is tall and not tall," and "#*x* is neither tall nor not tall," were "True," "False," or "Can't tell." This approach yielded responses that reproduced the indeterminacy about the borderline case, but the results were compatible with a gappy ("neither tall nor not tall" was true) and glutty ("tall and not tall" was true) interpretation, but there was a slight preference for gappy response. This difference was attributed to a pragmatic preference to deny a proposition directly rather than assert its negation. Finally, the hypothesis that vagueness is ignorance is rejected, because if epistemicism was true, then one would expect high levels of "Can't tell" responses in the borderline cases. This was not the case.

David Ripley (2011) reframes the problem as being squarely about what he calls "borderline contradictions," cases that might be either "gappy" or "glutty." He reports results of an experiment that was, in part, an attempt to control for the contextual variability appealed to in (Bonini et al. 1999) and in (Alxatib, Pelletier 2011). Ripley presented all participants with a visual prompt presenting a square and a circle with a decreasing distance between them, ordered from a maximum distance to a minimum distance, and asked to express agreement or disagreement with a prompt sentence for each case. Ripley found that responses that expressed stronger agreement with each of the borderline contradiction statements were most common where the circle and square were neither farthest apart nor closest together. Ripley takes this to be evidence that while the borderline phenomena are reproducible, the results are compatible with both contextual semantics using a classical logic and dialetheist semantics using a paraconsistent logic.

Paul Egré and Jérémy Zehr (2018) performed a series of studies where participants were presented with a description that was an example of a borderline case within a specified context. For example:

> A survey on wealth has been conducted in your country. In the population there are people with a very high degree of wealth and people with a very low degree of wealth. Then there are people who lie in the middle between these two areas.

> Imagine that Sam is one of the people in the middle range. Comparing Sam to other people in the population, is it true to say the following?

Participants were then invited to reply "Yes" or "No" to a set of sentences of the following sort:

> Sam is neither rich nor not rich.
> Sam is rich and not rich.
> Sam is richer than at least one other person.
> Sam is richer than everyone else.

The prompts were paired with a variant that used an antonym (e.g., "poor" for "not rich"), and the terms used were graded as being human related (rich, old, heavy, tall) or object related (fast, large, loud, wide).

Egré and Zehr claim to have confirmed both gappy and glutty borderline contradictions are employed by "native speakers" to describe borderline cases. They further report that for "gradable" adjectives, gappy contradictions were preferred to glutty contradictions.

While such descriptive findings provide fodder for various explanatory theories, they can also serve to inform examinations of the contours and scope of folk concepts. In the present case, the convergence in the empirical results suggests that there is a strong tendency among the folk to judge borderline cases of vagueness to be neither true nor false.

What we learn from such studies is not the mere fact of the variability of individual responses, nor the variability of various groupings, though these may be of philosophical interest by themselves (of an Identifying sort). Rather, such studies provide evidence for a convergence of sorts upon how categorization judgments work in philosophically salient cases, and whether these tendencies interact with familiar principles of rationality.

In the case of vagueness, we discover that the sorites emerges from ordinary categorization judgments employing a vague concept. If this very categorization act is not ruled out, then what we discover is that there is evidence against the view that our concepts may be precise or imprecise. This provides impetus for further philosophical speculation about the metaphysical import of vague or imprecise concepts — for instance, even if there is a property of being a square, is there a property of being roughly hexagonal?

With respect to the Identifying route, there is little doubt that it matters that the empirical results show that borderline cases are intrinsic to our understanding of vagueness, both psychologically and philosophically. The convergence of empirical results in support of the "gappy" rather than "glutty"

understanding of borderline cases can be used to support either a supervaluationist or epistemicist philosophical account of vagueness. Here, the bridge from psychological finding to philosophical evidence follows the Tracking route. Likewise, the same evidence might cause difficulty for defenders of approaches to vagueness rooted in non-classical logics — for instance, dialetheism. Such is the route of Pruning. It is also possible that the empirical evidence might be employed to revise our conception of what a vague concept is like. Perhaps, we might have understood borderline cases as tolerant (easily allowing a person to be both tall and not tall), whereas the findings support a picture of vague concepts as restrictive and where the last tall person in a series need not stand beside the first non-tall person. If so, then the revision of our thinking will help to reshape our thinking about categorization generally and might even release us from the intellectual labyrinth of the sorites paradox.

CONCLUSION: FROM PSYCHOLOGICAL FINDINGS TO PHILOSOPHICAL RESULTS

We take ourselves to have accomplished two main objectives here. First, to complement the traditional distinction between the positive and negative program, we have proposed a taxonomy for x-phi of routes that track inference and influence patterns, starting from psychological findings and arriving at philosophical results. The Identifying route holds that some psychological findings should be recognized as already of philosophical significance; the Tracking route holds that some psychological findings have *prima facie* evidential ramifications for philosophical questions; the Pruning route holds that some findings do not serve directly as evidence but do influence the total philosophical evidence set by defeating other pieces of first-order evidence; and the Sculpting route holds that learning facts about human psychology can have its own further consequences for human psychology itself and come to reshape or mold the judgments and inclinations that will get brought to bear in philosophical inquiry.

We are not committed to the success of any particular version of those routes, as applied to some particular body of findings and philosophical debate; rather, we present them instead as general inferential strategies that are at least apparently plausible and cogent, to display the wide and diverse range of options x-phi practitioners have in order to (with one last shift of metaphor) turn their empirical harvest into good philosophical bread. In-

deed, it is not just the experimentalists themselves who can make use of the items in this taxonomy.

We have furthermore illustrated the potential long reach of these routes by considering several distinct applications within philosophical logic, which might have antecedently been thought to be inhospitable terrain to the deployment of x-phi. To conclude on a conjectural note, let us suggest that the fourfold route can also be helpful in understanding a different kind of empirical fact: the development of x-phi itself.

Even if, in the past, there was much more continuity between philosophy and psychology, one might reasonably have thought that that had been brought to an end culminating in the mid-century fission of academic Departments of Philosophy and Psychology into two separate departments with distinct methods and research questions. That x-phi can be both experimental and philosophical is a metaphilosophically significant result, demonstrating that separation of departments does not entail that philosophy and psychology are best understood as sequestered and unrelated forms of inquiry. Even if institutionally disjoint,[33] methodologically their intersection is splendidly non-null.

This flourishing of x-phi over the last two decades, and the impact it has had on a range of traditional literatures, also serves as evidence in metaphilosophical debates about the nature of philosophical inquiry. Williamson poses the following question: "If mathematics is an armchair science, why not philosophy too?" (2007: 4) and seems in context to be expecting something of an "Indeed, why not?" to be elicited in the reader. Yet, we would like to echo the sentiments of Laurie Paul when she was attempting to show how x-phi could be a deep resource for analytic metaphysics:

> First, there was a general methodological point about theorizing. I argued that philosophy is increasingly concerned with making and defending claims about the world rather than claims about concepts, and that philosophical theories in this vein can be understood as classes of models that purport to describe and represent the world. Second, there was a more specific point about the role of experimental work with respect to this sort of philosophical theorizing. Here, I argued that ordinary judgments are used to build and evaluate the models of philosophical theories of the world, and such judgments must be appropriately generated if we can use the models to infer conclusions about the world. The need for a determination of whether and how ordinary judgments are appropriately generated provides a new role for empirical work in contemporary philosophy. (Paul 2010: 475)

---

[33] Notwithstanding the robust flourishing of interdisciplinary cognitive science centers, which frequently feature contributions from philosophy departments.

To the extent that even traditionally-minded philosophers must now often reckon with x-phi results — even in such domains as metaphysics and philosophical logic — seems to indicate that the answer to Williamson's question may in fact be more complicated than its rhetorical presentation would have suggested.

At the same time, many philosophers (perhaps including Williamson) have been happy to trust their armchair sense of what sort of thing philosophy is and how it ought to operate. One lesson from the flourishing of x-phi — and how unexpected it was at the time — is that such armchair metaphilosophical judgments should be taken now with much greater caution. Thus does x-phi offer the means to pursue a metaphilosophical inference along the third route.

Finally, we hold out hope that more philosophers do not treat x-phi merely as some sort of alien assault upon our special, plush-seated aerie, but rather as an ally who has come to the party bearing an abundance of varied conceptual and evidential resources. As this happens, and more and more of the community starts to absorb the findings and, furthermore, to gain comfort with the methods, then a version of the fourth route will apply here as well. That is, even beyond the arguments and evidence, the phenomena of x-phi itself may yield a subterranean and beneficial impact on our collective conception of what philosophy is and can be. Experimental philosophy's childhood was a bit rough, often scrapping with the bigger kids on the academic playground; changing metaphilosophical attitudes may bring it a more irenic adolescence.

BIBLIOGRAPHY

Adams F., Steadman A. (2004a), "Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding?," *Analysis* 64(2), 173-181. https://doi.org/10.1093/analys/64.2.173

Adams F., Steadman A. (2004b), "Intentional Actions and Moral Considerations: Still Pragmatic," *Analysis* 64(3), 268-276. https://doi.org/10.1093/analys/64.3.268

Alexander J. (2016), "Philosophical Expertise" [in:] *A Companion to Experimental Philosophy*, W. Buckwalter, J. Sytsma (eds.), Oxford: Wiley, 555-567. https://doi.org/10.1002/9781118661666.ch39

Alexander J., Weinberg J. (2006), "Analytic Epistemology and Experimental Philosophy," *Philosophy Compass* 2, 56-80. https://doi.org/10.1111/j.1747-9991.2006.00048.x

Alxatib S., Pelletier F. J. (2011), "The Psychology of Vagueness: Borderline Cases and Contradictions," *Mind and Language* 26, 287-326. https://doi.org/10.1111/j.1468-0017.2011.01419.x

Andow J. (2020), "Fully Experimental Conceptual Engineering," *Inquiry*, 1-27. https://doi.org/10.1080/0020174X.2020.1850339

Asay J. (2021), "Primitivism about Truth" [in:] *The Nature of Truth: Classic and Contemporary Perspectives*, N. Kellen, J. Kim, M. P. Lynch, J. Wyatt (eds.), 2nd ed., Cambridge, MA: MIT Press, 525-538.

Austin J. L. (1979), "A Plea for Excuses" [in:] *Philosophical Papers*, J. O. Urmson, G. J. Warnock (eds.), Oxford: Clarendon Press, 175-203. https://doi.org/10.1093/019283021X.003.0008

Balaguer M. (2016), "Conceptual Analysis and X-Phi," *Synthese* 193: 2367-2388. https://doi.org/10.1007/s11229-015-0848-4

Barnard R., Ulatowski J. (2016), "Tarski's 1944 Polemical Remarks and Næss' 'Experimental Philosophy'," *Erkenntnis* 81(4), 457-477. https://doi.org/10.1007/s10670-015-9750-1

Barnard R., Ulatowski J., Weinberg J. (2017), "Thinking about the Liar, Fast and Slow" [in:] *Reflections on the Liar*, B. Armour-Garb (ed.), Oxford: Oxford University Press, 39-70. https://doi.org/10.1093/oso/9780199896042.003.0003

Barwise J., Etchemendy J. (1987), *The Liar: An Essay on Truth and Circularity*, Oxford: Oxford University Press.

Baz A. (2018), *The Crisis of Method in Contemporary Analytic Philosophy*, Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198801887.001.0001

Black M. (ed.) (1950), *Philosophical Analysis: A Collection of Essays*, Englewood Cliffs, NJ: Prentice-Hall.

Block N., Stalnaker R. (1999), "Conceptual Analysis, Dualism, and the Explanatory Gap," *Philosophical Review* 108, 1-46. https://doi.org/10.2307/2998259

Bonini N., Osherson D., Viale R., Williamson T. (1999), "On the Psychology of Vague Predicates," *Mind and Language* 14, 377-393. https://doi.org/10.1111/1468-0017.00117

Buckland L., Lindauer M., Rodriguez-Arias D., Véliz C. (forthcoming), "Testing the Motivational Strength of Positive and Negative Duty Arguments Regarding Global Poverty," *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-021-00555-4

Buckwalter W. (2010), "Knowledge Isn't Closed on Saturday: A Study in Ordinary Language," *Review of Philosophy and Psychology* 1, 395-406. https://doi.org/10.1007/s13164-010-0030-3

Buttelmann D., Buttelmann F., Carpenter M., Call J., Tomasello M. (2017), "Great Apes Distinguish True from False Beliefs in an Interactive Helping Task," *PLoS ONE* 12(4): e0173793. https://doi.org/10.1371/journal.pone.0173793

Buttelmann D., Carpenter M., Tomasello M. (2009), "Eighteen-Month-Old Infants Show False Belief Understanding in an Active Helping Paradigm," *Cognition* 112, 337-343. https://doi.org/10.1016/j.cognition.2009.05.006

Carnap R. (1955), "Meaning and Synonymy in Natural Languages," *Philosophical Studies* 6(3): 33-47. https://doi.org/10.1007/BF02330951

Chapman S. (2011), "Arne Næss and Empirical Semantics," *Inquiry* 54(1), 18-30. https://doi.org/10.1080/0020174X.2011.542946

Chapman S. (2018), "The Experimental and the Empirical: Arne Næss' Statistical Approach to Philosophy," *British Journal for the History of Philosophy* 26, 961-981. https://doi.org/10.1080/09608788.2017.1336075

Cullen S. (2010), "Survey-Driven Romanticism," *Review of Philosophy and Psychology* 1(3): 275-296. https://doi.org/10.1007/s13164-009-0016-1

Cushman F. (2007), "The Effect of Moral Judgment on Causal and Intentional Attribution: What We Say, or How We Think?," unpublished manuscript, Harvard University.
DeBrigard F. (2010), "If You Like It, Does It Matter If It's Real?," *Philosophical Psychology* 23(1), 43-57. https://doi.org/10.1080/09515080903532290
Daniels N. (1979), "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *Journal of Philosophy* 76(5), 256-282. https://doi.org/10.2307/2025881
DeRose K. (2005), "The Ordinary Language Basis for Contextualism, and the New Invariantism," *Philosophical Quarterly* 55, 172-198. https://doi.org/10.1111/j.0031-8094.2005.00394.x
Deutsch M. (2015), *The Myth of the Intuitive*, Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262028950.001.0001
Egré P., Zehr J. (2018), "Are Gaps Preferred to Gluts? A Closer Look at Borderline Contradictions" [in:] *The Semantics of Gradability, Vagueness, and Scale Structure*, E. Castroviejo, L. McNally, G. Weidman Sassoon (eds.), Dordrecht: Springer, 25-58. https://doi.org/10.1007/978-3-319-77791-7_2
Feltz A., Zapertine C. (2010), "Do You Know More When It Matters Less?," *Philosophical Psychology* 23(5), 683-706. https://doi.org/10.1080/09515089.2010.514572
Fischer E. (2014), "Verbal Fallacies and Philosophical Intuitions: The Continuing Relevance of Ordinary Language Analysis" [in:] *Austin on Language*, B. Garvey (ed.), Basingstoke, UK: Palgrave Macmillan, 124-140. https://doi.org/10.1007/s11229-019-02081-4
Fischer E., Engelhardt P., Horvath J., Ohtani H. (2019), "Experimental Ordinary Language Philosophy: A Cross-linguistic Study of Defeasible Default Inferences," *Synthese* 198(2): 1029-1070. https://doi.org/10.1007/s11229-019-02081-4
Fraser A. (1891), "Visualization as a Chief Source of the Psychology of Hobbes, Locke, Berkeley, and Hume," *The American Journal of Psychology* 4, 230-247. https://doi.org/10.2307/1411269
Fraser A. (1893), "The Psychological Basis of Hegelism," *The American Journal of Psychology* 5, 472-495. https://doi.org/10.2307/1411914
Gendler T. (2000), "The Puzzle of Imaginative Resistance," *The Journal of Philosophy* 97, 55-81. https://doi.org/10.2307/2678446
Gerken M., Beebe J. R. (2016), "Knowledge In and Out of Contrast," *Noûs* 50, 133-164. https://doi.org/10.1111/nous.12064
Goldman A. (1979), "What is Justified Belief?" [in:] *Justification and Knowledge*, G. Pappas (ed.), The Hague: Reidel. https://doi.org/10.1007/978-94-009-9493-5_1
Gonnerman C., Ulatowski J., Sytsma J. (forthcoming), "The History and Philosophy of Experimental Philosophy" [in:] *Complete Compendium of Experimental Philosophy*, M. Bauer (ed.), New York: Continuum.
Goodman N. (1983), *Fact, Fiction, and Forecast*, 4th ed., Cambridge, MA: Harvard University Press.
Greene J. (2008), "The Secret Joke of Kant's Soul" [in:] *Moral Psychology,* vol. 3: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, W. Sinnott-Armstrong (ed.), Cambridge, MA: The MIT Press, 35-79.
Greene J. (2013), *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*, New York: Penguin.
Hansen N., Chemla E. (2015), "Linguistic Experiments and Ordinary Language Philosophy," *Ratio* 28(4), 422-445. https://doi.org/10.1111/rati.12112

Hansen N., Porter J. D., Francis K. (2019), "A Corpus Study of 'Know': On the Verification of Philosophers' Frequency Claims about Language," *Episteme*, 1-27. https://doi.org/10.1017/epi.2019.15

Hewitt S. (2010), "What Do Our Intuitions about the Experience Machine Really Tell Us about Hedonism?," *Philosophical Studies* 151(3), 331-349. https://doi.org/10.1007/s11098-009-9440-4

Horowitz T. (1998), "Philosophical Intuitions and Psychological Theory," *Ethics* 108, 367-385. https://doi.org/10.1086/233809

Horne Z., Livengood J. (2017), "Ordering Effects, Updating Effects, and the Specter of Global Skepticism," *Synthese* 194, 1189-1218. https://doi.org/10.1007/s11229-015-0985-9

Horschler D. J., MacLean E. L., Santos L. R. (2020), "Do Non-human Primates Really Represent Others' Beliefs?," *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.05.009

Horvath J. (2010), "How (Not) to React to Experimental Philosophy," *Philosophical Psychology* 23(4), 447-480. https://doi.org/10.1080/09515089.2010.505878

Ichikawa J. (2009), "Explaining Away Intuitions," *Studia Philosophica Estonica* 2, 94-116. https://doi.org/10.12697/spe.2009.2.2.06

Irikefe P. O. (2020), "A Fresh Look at the Expertise Reply to the Variation Problem," *Philosophical Psychology* 33, 1-28. https://doi.org/10.1080/09515089.2020.1761541

Jackson F. (1998), *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Oxford University Press.

Jago M. (2018), *What is Truth?*, Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198823810.001.0001

Kahneman D., Tversky A. (1982), "The Psychology of Preferences," *Scientific American* 246, 136-142. https://doi.org/10.1038/scientificamerican0182-160

Kelly D. (2011), *Yuck! The Nature and Moral Significance of Disgust*, Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/8303.001.0001

Kelly D. (2018), "Cast in a Bad Light or Reflected in a Dark Mirror? Cognitive Science and the Projecting Mind" [in:] *The Moral Psychology of Disgust*, N. Strohminger, V. Kumar (eds.), London: Routledge, 171-194.

Kim M., Yuan Y. (2015), "No Cross-cultural Differences in the Gettier Car Case Intuition: A Replication Study of Weinberg et al. 2001," *Episteme* 12(3): 355-361. https://doi.org/10.1017/epi.2015.17

Kneer M., Colaço D., Alexander J., Machery E. (forthcoming), "On Second Thought: Reflections on the Reflection Defense" [in:] *Oxford Studies in Experimental Philosophy*, J. Knobe, S. Nichols, T. Gendler (eds.), Oxford: Oxford University Press.

Knetsch J. L. (1989), "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *American Economic Review* 79, 1277-1284.

Knobe J. (2003). "Intentional Action and Side Effects in Ordinary Language," *Analysis* 63(3), 190-193. https://doi.org/10.1093/analys/63.3.190

Knobe J. (2004), "Intention, Intentional Action and Moral Considerations," *Analysis* 64(2), 181-187. https://doi.org/10.1093/analys/64.2.181

Knobe J. (2016), "Experimental Philosophy is Cognitive Science" [in:] *A Companion to Experimental Philosophy*, W. Buckwalter, J. Sytsma (eds.), Oxford: Wiley, 37-52. https://doi.org/10.1002/9781118661666.ch3

Koch S. (2019), "Carnapian Explications, Experimental Philosophy, and Fruitful Concepts," *Inquiry: An Interdisciplinary Journal of Philosophy* 62(6), 700-717. https://doi.org/10.1080/0020174X.2019.1567381

Koch S. (2020), "Engineering What? On Concepts in Conceptual Engineering," *Synthese* 199, 1955-1975. https://doi.org/10.1007/s11229-020-02868-w

Kripke S. (1980), *Naming and Necessity*, Cambridge: Harvard University Press.

Lillehammer H. (ed.) (forthcoming), *The Trolley Problem*, Cambridge: Cambridge University Press.

Lindauer M. (2020), "Experimental Philosophy and the Fruitfulness of Normative Concepts," *Philosophical Studies* 177(8), 2129-2152. https://doi.org/10.1007/s11098-019-01302-3

Machery E. (2008), "The Folk Concept of Intentional Action: Philosophical and Experimental Issues," *Mind and Language* 23(2), 165-189. https://doi.org/10.1111/j.1468-0017.2007.00336.x

Machery E., Mallon R., Nichols S., Stich S. P. (2004), "Semantics, Cross-Cultural Style," *Cognition* 92, B1-B12. https://doi.org/10.1016/j.cognition.2003.10.003

Machery E., Stich S., Rose D., Chatterjee A., Karasawa K., Struchiner N., Sirker S., Usui N., Hashimoto T. (2017), "Gettier across Cultures," *Noûs* 51(3), 645-664. https://doi.org/10.1111/nous.12110

Maudlin T. (2008), *Truth and Paradox: Solving the Riddles*, Oxford: Oxford University Press.

May J., Sinnott-Armstrong W., Hull J. G., Zimmerman A. (2010), "Practical Interests, Relevant Alternatives, and Knowledge Attributions: An Empirical Study," *Review of Philosophy and Psychology* 1(2), 265-273. https://doi.org/10.1007/s13164-009-0014-3

McKenna M., Pereboom D. (2014), *Free Will: A Contemporary Introduction*, London: Routledge.

Meskin A., Weinberg J. (2012), "Imagination Unblocked" [in:] *The Aesthetic Mind: Philosophy and Psychology*, E. Schellekens, P. Goldie (eds.), Oxford: Oxford University Press, 239-253. https://doi.org/10.1093/acprof:oso/9780199691517.003.0014

Moore J. (1939), Review of A. Næss: *"Truth" as Conceived by Those Who Are Not Professional Philosophers*, *The American Journal of Psychology* 52, 489-490. https://doi.org/10.2307/1416771

Nadelhoffer T., Nahmias E., Morris S., Turner J. (2004). "The Phenomenology of Free Will," *Journal of Consciousness Studies* 11(7-8), 162-179.

Nadelhoffer T., Nahmias E. (2007), "The Past and Future of Experimental Philosophy," *Philosophical Explorations* 10, 123-149. https://doi.org/10.1080/13869790701305921

Nadelhoffer T., Shepard J., Crone D. L., Everett J. A. C., Earp B. D., Levy N. (2020), "Does Encouraging a Belief in Determinism Increase Cheating? Reconsidering the Value of Believing in Free Will," *Cognition* 203, 1-13. https://doi.org/10.1016/j.cognition.2020.104342

Nado J. (2014), "Philosophical Expertise," *Philosophy Compass* 9, 631-641. https://doi.org/10.1111/phc3.12154

Nado J. (2015), "Philosophical Expertise and Scientific Expertise," *Philosophical Psychology* 28, 1026-1044. https://doi.org/10.1080/09515089.2014.961186

Nado J. (2017), "How To Think About Philosophical Methodology," *Journal of Indian Council of Philosophical Research* 34, 447-463. https://doi.org/10.1007/s40961-017-0116-8

Nado J. (2021), "Conceptual Engineering via Experimental Philosophy," *Inquiry: An Interdisciplinary Journal of Philosophy* 64(1-2), 76-96. https://doi.org/10.1080/0020174X.2019.1667870

Næss A. (1938a), *"Truth" as Conceived by Those Who Are Not Professional Philosophers (Skrifter Utgitt av Det Norske Videnskaps-Akademi I Oslo Il. Hist.-Filos.* Klass 1938 No. 4*)*, Oslo: I Komisjon Hos Jacob Dybwad.

Næss A. (1938b), "Common-Sense and Truth," *Theoria* 4, 39-58. https://doi.org/10.1111/j.1755-2567.1938.tb00438.x

Næss A. (1953), "Toward a Theory of Interpretation and Preciseness," *Theoria* 15, 220-241. https://doi.org/10.1111/j.1755-2567.1949.tb00152.x

Nagel E. (1939), Review of: A. Næss, *"Truth" as Conceived by Those Who Are Not Professional Philosophers*, *Journal of Philosophy* 36, 78-80. https://doi.org/10.2307/2017766

Nahmias E., Morris S., Nadelhoffer T., Turner J. (2006), "Is Incompatibilism Intuitive?," *Philosophy and Phenomenological Research* 73(1), 28-53. https://doi.org/10.1111/j.1933-1592.2006.tb00603.x

Nichols S., Knobe J. (2007), "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions," *Noûs* 41(4), 663-685. https://doi.org/10.1111/j.1468-0068.2007.00666.x

Nichols S., Roskies A. (2008), "Bringing Moral Responsibility Down to Earth," *The Journal of Philosophy* 105(7), 371-388. https://doi.org/10.5840/jphil2008105737

Nichols S., Ulatowski J. (2007), "Intuitions and Individual Differences: The Knobe Effect Revisited," *Mind and Language* 22(3), 346-365. https://doi.org/10.1111/j.1468-0017.2007.00312.x

Nozick R. (1974), *Anarchy, State, and Utopia*, New York: Basic Books.

Nozick R. (1989), *The Examined Life*, New York: Touchstone Books.

Paprzycka-Hausman K. (2015), "The Omissions Account of the Knobe Effect and the Asymmetry Challenge," *Mind and Language* 30(5), 550-571. https://doi.org/10.1111/mila.12090

Paul L. (2010), "A New Role for Experimental Work in Metaphysics," *Review of Philosophy and Psychology* 1, 461-476. https://doi.org/10.1007/s13164-010-0034-z

Pinder M. (2017a), "The Explication Defence of Arguments from Reference," *Erkenntnis* 82(6), 1253-1276. https://doi.org/10.1007/s10670-016-9868-9

Pinder M. (2017b), "Does Experimental Philosophy Have a Role to Play in Carnapian Explication?," *Ratio* 30(4), 443-461. https://doi.org/10.1111/rati.12164

Pinillos N. Á., Smith N., Nair G. S., Marchetto P., Mun C. (2011), "Philosophy's New Challenge: Experiments and Intentional Action," *Mind and Language* 26, 115-139. https://doi.org/10.1111/j.1468-0017.2010.01412.x

Plakias A. (2015), "Experimental Philosophy," *Oxford Handbook Online*, 1-18. https://doi.org/10.1093/oxfordhb/9780199935314.013.17

Priest G. (1997), "Sylvan's Box: A Short Story and Ten Morals," *Notre Dame Journal of Formal Logic* 38, 573-582. https://doi.org/10.1305/ndjfl/1039540770

Rawls J. (1971), *A Theory of Justice*, Oxford: Oxford University Press.

Rini R. (2015), "How Not to Test for Expertise," *Synthese* 192, 431-452. https://doi.org/10.1007/s11229-014-0579-y

Ripley D. (2011), "Contradictions at the Borders" [in:] *Vagueness in Communication,* R. Nouwen, R. an Rooij, U. Sauerland, H-C. Schmitz (eds.), Dordrecht: Springer, 169-188. https://doi.org/10.1007/978-3-642-18446-8_10

Ripley D. (2016), "Experimental Philosophical Logic" [in:] *A Companion to Experimental Philosophy*, W. Buckwalter, J. Sytsma (eds.), Oxford: Wiley, 523-534. https://doi.org/10.1002/9781118661666.ch36

Rose D. (2015), "Belief is Prior to Knowledge," *Episteme* 12(3), 385-399. https://doi.org/10.1017/epi.2015.21

Rowland R. (2017), "Our Intuitions about the Experience Machine," *Journal of Ethics and Social Philosophy* 12, 110-117. https://doi.org/10.26556/jesp.v12i1.216

Russell B. (1923), "Vagueness," *Australasian Journal of Philosophy* 1, 84-92. https://doi.org/10.1080/00048402308540623

Samuelson W., Zeckhauser R. (1988), "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty* 1(1), 7-59. https://doi.org/10.1007/BF00055564

Schubach J. (2017), "Explanatory Explication," *Philosophy and Phenomenological Research* 94(3), 672-710. https://doi.org/10.1111/phpr.12207

Schubach J. (2018), "Robustness Analysis as Explanatory Reasoning," *British Journal for the Philosophy of Science* 69, 275-300. https://doi.org/10.1093/bjps/axw008

Schwitzgebel E. (2009), "Do Ethicists Steal More Books?," *Philosophical Psychology* 22, 711-725. https://doi.org/10.1080/09515080903409952

Schwitzgebel E., Cokelet B., Singer P. (forthcoming), "Do Ethics Classes Influence Student Behavior? Case Study: Teaching the Ethics of Eating Meat," *Cognition*.

Schwitzgebel E., Rust J. (2012), "The Behavior of Ethicists" [in:] *A Companion to Experimental Philosophy*, W. Buckwalter, J. Sytsma (eds.), Oxford: Wiley, 225-233. https://doi.org/10.1002/9781118661666.ch15

Schwitzgebel E., Rust J., Huang L., Moore A., Coates J. (2012), "Ethicists' Courtesy at Philosophy Conferences," *Philosophical Psychology* 35, 331-340. https://doi.org/10.1080/09515089.2011.580524

Shearman A. T. (1907), "Intuition," *Proceedings of the Aristotelian Society* 7, 158-197. https://doi.org/10.1093/aristotelian/7.1.158

Shepherd J., Justus J. (2015), "X-Phi and Carnapian Explication," *Erkenntnis* 80(2), 381-402. https://doi.org/10.1007/s10670-014-9648-3

Silverstein M. (2000), "In Defense of Happiness: A Response to the Experience Machine," *Social Theory and Practice* 26(3), 279-300. https://doi.org/10.5840/soctheorpract200026225

Singer P. (1972), "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1, 229-243.

Sorrell T. (2017), "Experimental Philosophy and the History of Philosophy," *British Journal for the History of Philosophy* 26(5): 829-849. https://doi.org/10.1080/09608788.2017.1320971

Sripada C. S., Stanley J. (2012), "Empirical Tests of Interest-Relative Invariantism," *Episteme* 9(1), 3-26. https://doi.org/10.1017/epi.2011.2

Stewart J. A. (1876), "Psychology: A Science or a Method?," *Mind* 1, 445-451. https://doi.org/10.1093/mind/os-1.4.445

Stich S. (1988), "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity," *Synthese* 74, 391-413. https://doi.org/10.1007/BF00869637

Stotz K., Griffiths P. E., Knight R. (2004), "How Biologists Conceptualize Genes: An Empirical Study," *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 35, 647-673. https://doi.org/10.1016/j.shpsc.2004.09.005

Sytsma J. (2010), "The Proper Province of Philosophy: Conceptual Analysis and Empirical Investigation," *Review of Philosophy and Psychology* 1(3), 427-445. https://doi.org/10.1007/s13164-010-0032-1

Sytsma J., Livengood J. (2012), "Experimental Philosophy and Philosophical Disputes," *Essays in Philosophy* 13(1), 145-161. https://doi.org/10.5840/eip20121319

Sytsma J., Livengood J. (2019), "On Experimental Philosophy and the History of Philosophy: A Reply to Sorrell," *British Journal for the History of Philosophy* 27(3): 635-647. https://doi.org/10.1080/09608788.2018.1539946

Tarski A. (1933/1983), "The Concept of Truth in Formalized Languages" [in:] *Logic, Semantics, Metamathematics*, J. Woodger (ed.), Indianapolis, IN: Hackett Publishing, 152-278.

Tarski A. (1944), "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* 4, 341-375. https://doi.org/10.2307/2102968

Ulatowski J. (2016), "Ordinary Truth in Tarski and Næss" [in:] *Uncovering Facts and Values*, J. Odrowąż-Sypniewska, A. Kuźniar (eds.), Leiden: Brill, 67-90. https://doi.org/10.1163/9789004312654_006

Ulatowski J. (2017), *Commonsense Pluralism about Truth: An Empirical Defence*, Basingstoke: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-69465-8

Ulatowski J. (2018), "Is There a Commonsense Conception of Truth?," *Philosophia* 46(4), 487-500. https://doi.org/10.1007/s11406-017-9941-x

Ulatowski J. (ms.), "Canberra Planning in Norwegian Wood: Experimental Work at the Antipodes," University of Waikato.

Urmson J. O., Quine W. V. O., and Hampshire S. (1969), "A Symposium on Austin's Method" [in:] *Symposium on J. L. Austin*, K. T. Fann (ed.), London: Routledge, 76-97.

Vohs K. D., Schooler J. W. (2008), "The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating," *Psychological Science* 19, 49-54. https://doi.org/10.1111/j.1467-9280.2008.02045.x

Warnock G. (1969), "John Langshaw Austin: A Biographical Sketch" [in:] *Symposium on J. L. Austin*, K. T. Fann (ed.), London: Routledge, 3-21.

Weijers D. (2014), "Nozick's Experience Machine Is Dead, Long Live the Experience Machine," *Philosophical Psychology* 27(4), 513-535. https://doi.org/10.1080/09515089.2012.757889

Weinberg J. M. (2016a), "Intuitions" [in:] *Oxford Handbook of Philosophical Methodology*, H. Cappelen, T. Szabo Gendler, J. Hawthorne (eds.), Oxford: Oxford University Press, 287-308. https://doi.org/10.1093/oxfordhb/9780199668779.013.25

Weinberg J. M. (2016b), "Going Positive by Going Negative: On Keeping X-Phi Relevant and Dangerous" [in:] *A Companion to Experimental Philosophy*, W. Buckwalter, J. Sytsma (eds.), Oxford: Wiley, 71-86. https://doi.org/10.1002/9781118661666.ch5

Weinberg J. M., Alexander J., Gonnerman C., Reuter S. (2012), "Restrictionism and Reflection: Challenge Deflected, or Simply Redirected?," *The Monist* 95, 200-222. https://doi.org/10.5840/monist201295212

Weinberg J., Nichols S., Stich S. (2001), "Normativity and Epistemic Intuitions," *Philosophical Topics* 29, 429-460. https://doi.org/10.5840/philtopics2001291/217

Werner K. (2020), "Philosophical Intuition Is the Capacity to Recognize One's Epistemic Position: An Old-Fashion Approach Based on Russell, Carnap, Wittgenstein, and Husserl," *Philosophia* 48(1): 1725-1751. https://doi.org/10.1007/s11406-020-00195-5

Williamson T. (1994), *Vagueness*, London: Routledge.
Williamson T. (2000), *Knowledge and Its Limits*, Oxford: Oxford University Press.
Williamson T. (2005), "Armchair Philosophy, Metaphysical Modality, and Counterfactual Thinking," *Proceedings of the Aristotelian Society* 105, 1-23. https://doi.org/10.1111/j.0066-7373.2004.00100.x
Williamson T. (2007), *The Philosophy of Philosophy*, Oxford: Wiley. https://doi.org/10.1002/9780470696675
Woolfolk R. (2013), "Experimental Philosophy: A Methodological Critique," *Metaphilosophy* 44(1-2): 79-87. https://doi.org/10.1111/meta.12016
Żuradzki T., Wiśniowska K. (2020), "A Data-Driven Argument in Bioethics: Why Theologically Grounded Concepts May Not Provide the Necessary Intellectual Resources to Discuss Inequality and Injustice in Healthcare Contexts," *American Journal of Bioethics* 20(12): 25-28. https://doi.org/10.1080/15265161.2020.1832617