

DISCIPLINE FILOSOFICHE

Anno XXV, numero 1, 2015

Quodlibet

DISCIPLINE FILOSOFICHE

Anno XXV, numero 1, 2015

Rivista fondata da Enzo Melandri.

Periodicità semestrale.

Aut. Tribunale di Macerata, n. 527/

stampa del 16. 12. 2005

ISSN: 1591-9625



Dipartimento di Filosofia e
Comunicazione

Direttore Stefano Besoli

Direttore responsabile Barnaba Maj

Comitato scientifico

Massimo Barale (Università di Pisa) †, Jocelyn Benoist (Université Paris-I Panthéon-Sorbonne), Giuseppe Cantillo (Università Federico II di Napoli), Jean-François Courtine (Université Paris-IV Sorbonne), Françoise Dastur (Université de Nice Sophia Antipolis), Roberta De Monticelli (Università Vita-Salute San Raffaele di Milano), Bianca Maria d'Ippolito (Università di Salerno) †, Massimo Ferrari (Università di Torino), Gottfried Gabriel (Friedrich Schiller Universität Jena), Gianna Gigliotti (Università Tor Vergata di Roma), Wolfhart Henckmann (Ludwig-Maximilians-Universität München), Douglas Hofstadter (Indiana University), John Lachs (Vanderbilt University), Claudio La Rocca (Università di Genova), Eugenio Mazzarella (Università Federico II di Napoli), Ernst Wolfgang Orth (Universität Trier), Renato Pettoello (Università di Milano), Manfred Sommer (Christian-Albrechts-Universität Kiel), Jürgen Stolzenberg (Martin-Luther-Universität Halle-Wittenberg), Francesco Saverio Trincia (Università La Sapienza di Roma), Frédéric Worms (École normale supérieure – ENS, Paris)

Comitato di redazione

Simona Bertolini, Francesco Bianchini, Roberto Brigati, Vincenzo Costa, Roberto Frega, Sebastiano Galanti Grollo, Michele Gardini, Alberto Gualandi, Luca Guidetti, Barnaba Maj, Giuliana Mancuso, Marina Manotta, Emanuele Mariani, Riccardo Martinelli, Maurizio Matteuzzi, Giovanni Matteucci, Venanzio Raspa, Alessandro Salice, Luca Vanzago, Giorgio Volpe

Direzione e redazione

Dipartimento di Filosofia e Comunicazione, via Zamboni 38 - 40126 Bologna.

Tel 051-2098344. Fax 051-2098355

E-mail Redazione: l.guidetti@unibo.it

Sito web: www.disciplinefilosofiche.it

Copertina: Augusto Wirbel

© Copyright 2015 Quodlibet

ISBN: 9788874627998

Quodlibet edizioni, via Santa Maria della Porta, 43, 62100 Macerata

tel. 0733-264965 fax 0733-267358

www.quodlibet.it; e-mail: ordini@quodlibet.it

Finito di stampare nel mese di novembre 2015 dalla Grafica Editrice Romana s.r.l., Roma.

Questa Rivista è stata pubblicata con un contributo di fondi R.F.O. dell'Ateneo di Bologna.

I saggi che compaiono in questa rivista sono sottoposti a double-blind peer-review.

Jonathan M. Weinberg

The Methodological Necessity of Experimental Philosophy

Abstract

Must philosophers incorporate tools of experimental science into their methodological toolbox? I argue here that they must. Tallying up all the resources that are now part of standard practice in analytic philosophy, we see the problem that they do not include adequate resources for detecting and correcting for their own biases and proclivities towards error. Methodologically sufficient resources for error-detection and error-correction can only come, in part, from the deployment of specific methods from the sciences. However, we need not imagine that the resulting methodological norms will be so empirically demanding as to require that all appeals to intuition must first be precertified by a thorough vetting by teams of scientists. Rather, I sketch a set of more moderate methodological norms for how we might best include these necessary tools of experimental philosophy.

Keywords: Experimental philosophy, Armchair philosophy, Philosophical methodology, Intuitions, Philosophical expertise, Epistemic demandingness.

1. *What is an armchair, that one might want to sit in it?*

About a decade and a half after papers started being published under the flag of “experimental philosophy”, it seems to me that most philosophers who have a view about such work, think that it can perhaps be of at least some modest benefit to the profession, or that it is, at worst, a bit of a distraction.¹ But my impression is also that most philosophers think that experimental philosophy is not *especially* relevant to what they do, and is something that one can take or leave as one chooses, perhaps according to one’s metaphilosophical tastes. I want to push the more ambitious line, however, that experimental philosophy (henceforth “x-phi”, according to recent usage) is in fact a necessary addition to our field’s methodological resources. I will argue that here by contending that our armchair resources – in a sense to be expanded on shortly – are too impoverished to satisfy the

¹ Although there are some important dissenters, e.g., Deutsch (2010), Cappelen (2012), I will not be engaging with them here. See my (2014) for a brief response.

needs of philosophical inquiry on the whole, and that moreover x-phi's tools can turn many of the sorts of screws that our armchair tools cannot touch.

Let us start, then, by considering the question: just what *can* be done from the armchair? I will be construing armchairhood both broadly and generously here, to try to capture the resources commonly deployed in current analytic philosophical practice. We can start with Timothy Williamson's gloss of armchair methods as follows:

Every armchair pursuit raises the question of whether its methods are adequate to its aims. The traditional methods of philosophy are armchair ones: they consist of thinking, without any special interaction with the world beyond the chair, such as measurement, observation or experiment would involve. (Williamson 2007, p. 1)

This seems a good start at capturing what philosophers mean by working from the armchair, but I think we can unpack a bit further.² Contemplating what sorts of resources do seem to get drawn on regularly across a wide range of the sorts of (mostly analytic) philosophy today that might be considered to be operating from the armchair, one can see that the resources generally taken to be fair game include at least the following:

- common sense and similar sorts of facts available to informal observation, both perceptual and intellectual (often, but not always, under the term “intuition”);
- the received general knowledge of the college-educated population, including even some fairly sophisticated scientific results so long as they are sufficiently well-entrenched at this point, such as the general outlines of modern physics and evolutionary biology. As a rough rule of thumb: it's the kind of scientific results that don't require you to offer any citations on their behalf when you appeal to them, or just by referencing a name (e.g., “Darwin”) or a title (e.g., “Special Relativity”) without more specific reference.

We should also add the following that are standard components of philosophical training:

- the history of philosophy, including the track records of various techniques and approaches, as well as an ample stockpile of potentially useful distinctions and technical terms;
- all of mathematics and formal logic, as needed;
- a well-elaborated theory of *argumentation*, including a fairly well-theorized set of norms both positive (e.g., select premises that plausibly will be granted by an opponent; make clear how your premises collectively necessitate your conclusion) and negative (e.g., don't argue in a circle, don't confuse use and men-

² What follows should be taken as generally of a piece with the sort of “a posteriori armchair” defended recently in Nolan (2015).

tion). This also includes norms for responding to arguments (e.g., what constitutes a successful counterexample).

Now, in many areas of philosophy, it is clear that even this ample set of resources does not exhaust what is commonly taken to be legitimate to draw upon. The philosophies of the specific sciences, to take an obvious example, will of course call upon a much more extensive and fine-grained mastery of current scientific results and controversies. Williamson acknowledges this observation about current philosophical methodology as well, when he “raises no objections to the idea that the results of scientific experiments are sometimes directly relevant to philosophical questions: for example, concerning the philosophy of time” (Williamson 2007, p. 6). Or consider the philosophies of the arts, in which we frequently see very sophisticated appeals to and engagements with matters like art history, music or film theory, or the technical practices of performers. It is not controversial that a topic-specific “philosophy of X” must engage with the specific contours of X. While these resources serve as key enrichments to the armchair resources enumerated above, they are not used in quite the same way. Typically (albeit not exclusively) the philosopher is purely a consumer of these resources, and is not adding back to them; and it is highly rare (though, again, not unheard of) for philosophers to draw on these resources outside of the philosophy of X. For example, contemporary physics shows up a lot in philosophy of physics (of course), and a bit in some particular debates in metaphysics, and hardly anywhere else. So, although none of this kinds of work seems legitimately called *armchair* philosophy, nonetheless we can still see even here a kind of *domain-specific extensions* to the armchair.

Let’s call the above picture of the analytic philosophical toolbox the *current analytic methodological consensus* (CAMC). Experimental philosophy looks to upset that consensus, in large part by incorporating not just select results but also *methods* from the sciences, especially the social sciences, directly into the core tools, and not as a domain-specific extension. Now, that’s reason enough for folks to be upset by it – no consensus is ever disturbed without its being disturbing – but I worry that there is a lot of confusion around the profession as to both the nature of the intended disruption, and the motives behind it. I aim to clarify both here, and in doing so, hopefully make clearer just what the nature of x-phi’s challenge to CAMC really is, and why even fairly traditionally-minded philosophers should perhaps look to embrace it nonetheless.

Before turning to that challenge, though, I want to note that CAMC does already admit of one kind of experimental philosophy as unproblematic, as offering no such disturbance of the methodological status quo. I think the profession on the whole is used to the idea that where philosophy bor-

ders other disciplines – and there are many such borders – there will be much good, constructive work that fully inhabits both sides of the disciplinary divide. There is excellent work that is both philosophy of language and linguistics, both history of philosophy (as part of philosophy more generally) and history of ideas (as part of history more generally), both philosophy of physics and the very edge of theoretical physics itself. Some experimental philosophy work is intended to contribute to the philosophy of psychology by adding to our knowledge of philosophically-interesting pieces of scientific psychology. Joshua Knobe is a prominent champion of this particular variety of x-phi (forthcoming). The work is not meant to engage with standing debates about knowledge or intention or causation as such, but rather to help us understand how human minds engage with these notions. We can understand Knobe-style x-phi as an instance of the CAMC, in which not just psychology's results but also psychology's tools are recruited, as a legitimate domain-specific extension.³

But when philosophers are not doing cognitive science or the philosophy of psychology (or the psychology of philosophy), of what use or relevance is x-phi to them? To the extent that x-phi offers only a contribution to one specific subfield of philosophy, it also to that extent may often be legitimately ignored by those doing work outside that subfield. It *can* be recruited when relevant, as philosophers often recruit useful ideas across sub-field boundaries (e.g., when epistemologists debating contextualism redeploy machinery originally from the philosophy of language). Yet it would not *need* not do so. X-phi-understood-as-cog-sci is of no more general methodological relevance to other parts of philosophy than, say, current methods of physics are outside of contemporary philosophy of physics.

To bring into view the more general relevance of x-phi, beyond its being a contribution to cognitive science, let me springboard off of a recent admonition for methodological reflection from Williamson. While his *Philosophy of Philosophy* opens with the gloss on the armchair we adverted to above, towards the end of that book, he urges armchair philosophers that they “must do better” by attending closely to the linguistic aspects of our philosophical activities, on a model of how scientists must understand the tools that they deploy in their investigations:

³ There is a tangled issue here regarding the relationship between experimental philosophy and naturalistic philosophy, especially where philosophy-of-X and highly-theoretical-X shade into each other, as happens with some frequency in the philosophy of physics and the philosophy of cognitive science (see, e.g., Prinz 2007). I don't take the philosophers in these areas to be the target of my paper here; I expect they would consider themselves to be operating outside the armchair already.

Philosophers who refuse to bother about semantics, on the grounds that they want to study the non-linguistic world, not our talk about the world, resemble scientists who refuse to bother about the theory of their instruments, on the grounds that they want to study the world, not our observation of it. Such an attitude may be good enough for amateurs; applied to more advanced inquiries, it produces crude errors. Those metaphysicians who ignore language in order not to project it onto the world are the very ones most likely to fall into just that fallacy, because their carelessness with the structure of the language in which they reason makes them insensitive to subtle differences between valid and invalid reasoning. (Williamson 2007, pp. 284-285)

Perhaps optics is not part of astronomy proper; nonetheless, a decent astronomer had better know a lot about how light interacts with lenses, mirrors, and the atmosphere (or radio waves, and so on). And thus the experimentalist can argue on a closely parallel line that, even if x-phi is fundamentally psychological, its philosophical relevance will extend far beyond cognitive science and the philosophy of psychology itself:

Philosophers who refuse to bother about the empirically-discoverable workings of our minds, on the grounds that they want to study the extramental world, not our thought or concepts about that world, resemble scientists who refuse to bother about the theory of their instruments on the grounds that they want to study the world, not our observation of it. Such an attitude may be good enough for amateurs; applied to more advanced inquiries, it produces crude errors. Those metaphysicians who ignore the empirical in order to preserve the ideal of methodological self-sufficiency are the very ones most likely to fall into error, because their carelessness of the structure of the human mind with which they reason makes them insensitive to subtle differences between accurate and inaccurate observations.⁴

The danger here is not just one of the *possibility* of error – we are surely already aware of all sorts of ways in which philosopher can and do make mistakes, and we didn't need x-phi to teach us merely that philosophers are fallible – but, rather, the threat of stumbling into *unnoticed and heretofore unnoticeable* pitfalls, ones invisible to our current methodological resources. For what CAMC *can't* do, even when we include its domain-specific extensions, is *sufficiently detect its own susceptibilities to bias and error*. To be clear, this is not an across-the-board problem for the CAMC. In particular, I have no evidence to offer that would cast doubt on the formal sciences' current adequacy to detecting its own threats of error. For over the centuries those methods have developed elaborate and sophisticated practices of formalization, of laying bare one's axioms and rules of inference and the like, of articulated proofs and indeed the rigorous checking of such proofs.

⁴ I am drawing liberally here from my (2009).

I similarly have no doubts to raise about methods within the history of philosophy, with its practices of scholarship and archival work, for example.

2. *Experimental philosophy and the challenge of inappropriate sensitivity*

My concerns are primarily about the first item on my ledger of CAMC resources: what are the deficiencies in our ordinary capacities of ordinary observation and intuition, and common-sense generalization, such that x -phi can at least in principle improve on them?

Proponents of this challenge to the methodological self-sufficiency of the armchair often point to vectors of *inappropriate sensitivity*. We want the deliverances of these capacities to track whatever really does make a difference between, say, knowing and not-knowing, or free actions and unfree, but at the same time we do not want them to be driven by factors outside of the relevant philosophical truths. In general, factors like demographics, order of presentation, subtle and philosophically-irrelevant shifts in wording, or even the font that a case is presented in – these are all factors that, while they do not seem likely to be good candidates for inclusion in our best theories of knowledge, agency, moral goodness and the like, they are all nonetheless factors for which there is growing evidence that our intuitions are problematically sensitive.⁵

Let me be clear about what the challenge of inappropriate sensitivity is *not*. First, as noted above but worth emphasizing, it is not the same as mere fallibility – and the methods advocated in experimental philosophy are not themselves infallible, after all. Any proposed piece of methodological advice, “don’t trust any fallible sources” will surely run into both a wildly over-generalized skepticism (since near enough to all human epistemic resources are fallible), as well as self-defeat (since the source of that advice will surely itself be fallible).

Relatedly, the challenge of inappropriate sensitivity does not require imposing the hyperbolic, skeptical requirement that all methodological resources be non-circularly calibrated, as e.g. some have worried is the case with Cummins (1998). The problem is not that the resources of CAMC merely lack for some sort of independent certification, while otherwise perhaps being perfectly fine; the problem is that we have actual positive reason to think that they have flaws that are beyond their collective ability to correct.

⁵ See Buckwalter et al. (2012) for a number of instances of such results. I should note that, while some have not replicated well in the interim, such as my own (2001), many other results have been replicated successfully and indeed extended, such as those of Machery et al. (2004), Swain et al. (2008), and Feltz and Cokely (2009).

Moreover, it is not any sort of self-hating philosopher's appeal to a blinkered scientism. The argument isn't, "CAMC does not include some methodological characteristic necessary to count as science, and for that reason it is inadequate". It is not a matter of imposing an alien methodological criterion onto philosophy, one that philosophy has no reason of its own to endorse. The challengers presume that successfully tracking the truth about matters philosophical is a value that is internal to philosophy, and indeed central to many forms of it.⁶ The reason that experimental philosophers agitate for a larger incorporation of scientific methods into philosophy at large is not simply – not at all – because they qualify for some magic status of "science!". We take ourselves instead to have good reason to think that those methods are the best available to address the specific deficiencies observed within CAMC. I will illustrate this point in terms of two large classes of such deficiencies: cognitive diversity, and subtle contextual effects.

Cognitive diversity will be obscured by our own natural sampling of those who are like us, and self-selection within the profession. There is the famous story of the film critic Pauline Kael, acknowledging the biased sample of her own professional world in the context of the 1972 election: "I live in a rather special world. I only know one person who voted for Nixon. Where they are I don't know. They're outside my ken. But sometimes when I'm in a theater I can feel them". In the context of a presidential election, all of us can get good feedback as to how far our own local communities may diverge from the larger body politic, and Kael may well have been more sensitive to such divergences than your typical intellectual, due to that time in the theaters, and to her copious acute attention to the popular cinema. In contrast, your typical philosopher, considering a standard (and, standardly, at least a little bit weird) thought-experiment, will likely be more at sea as to whether they are or are not on the same wavelength as any larger community. It is easy for most of us to be out of tune with the folk in general, despite interactions such as those in the classroom that may give us the illusion of receiving adequate feedback (Stich and Weinberg 2001). But for that matter, it is not that hard for one sub-community of philosophers to get itself out of sync with the rest of the profession. Anecdotally, it seems to me that some famous thought-experiments elicit a much wider array of responses even in the profession than their original authors may have suspected, such as Swampman, the fake barn, and high stakes/low stakes bank

⁶ However, those whose metaphilosophies that do not traffic in, or even oppose, thinking of philosophy in such terms will also rightly not find much of relevance to them in experimental philosophy's methodological challenge.

cases. It would be good to be able to get beyond anecdotes, though – between-philosopher variation would be a good direction for future research!

And subtle contextual effects like framing, order, font choice, and so on are, well, subtle, and operate largely unconsciously. They are thus invisible to introspection, and not likely to be revealed to unaided, unsystematic observation. They are just the sort of thing that it took experimental psychology to uncover in the first place, after all. Even should some philosopher notice such effects, they will often lack a clear enough evidential basis to persuade the profession of it more widely, and at best it will remain a debated and debatable point, as so many attempted “explainings away” of unwanted intuitions remain, in the literature. And we should expect that many armchair conjectures as to the underlying causal working of these case verdicts will be mistaken, as CAMC just does not have the kind of resolving power to separate subtle effects that are really there from those that might be merely a mistaken conjecture on the part of a theorist.⁷

So these two general kinds of error vectors will by and large lie beyond the power of CAMC to detect and correct. Yet scientific methods can overcome these problems largely by looking for them directly. We can design studies to sample deliberately across a broad range of participants, and any hypothesized differences can be looked for directly. Many subtle effects can be controlled for by good experimental design as well. For example, order effects can be controlled for by presenting sets of cases in different orders to different participants, as is fairly standard scientific practice. Moreover, these methods allow us to use statistics in order to help pick out real effects from illusory ones.

A further complication here is that the armchair can drastically misreckon its own degree of competence. Precisely because we are susceptible to all sorts of biases and errors that we typically cannot detect using only armchair resources, we will tend to overestimate our capacity for detecting and correcting for errors. Biases detected and corrected for will count positively in our estimate of that capacity, but those that are not detected in the first place will, for that very reason, not be able to figure into that evaluation.

A nice illustration of this problem comes from the literature debating x -phi itself. A number of philosophers have claimed that, while the sort of undergraduate or otherwise non-specialist subjects in x -phi studies may display a diverse set of responses or be susceptible to funny sorts of unconscious effects, we should nonetheless expect *expert philosophers* to display much greater uniformity and immunity to such effects (Ludwig 2007; Hales 2006;

⁷ See Ichikawa (2009) for an exploration of some of the issues around the topic of explaining away intuitions.

Williamson 2005). Where such a line might be offered merely as a hypothesis for consideration and investigation, I would have no objections to it. But many have presented it as a claim so clearly true, that at a minimum it pushes the burden of proof all the way over on to the would-be critic of CAMC. And yet this claim of philosophical expertise turns out, for starters, to be not particularly consistent with the general scientific findings on expertise (Weinberg et al. 2010). And in fact, a growing set of results looking specifically at whether philosophers are immune to the threats of diversity and unconscious biases have by and large disconfirmed the key claims of this expertise defense.⁸ I expect that there will be some zones where philosophical training does prove to screen off some of these error vectors. But we will only be able to uncover them by the careful application of the methods of the empirical sciences – that is, by abandoning CAMC and doing a fair amount of x-phi.

Now, some readers might have been impatient since section 1 to object that I have left out some key resources that belong to CAMC. And it would not surprise me if that were so. Any such unfortunate omissions on my part should, however, be measured according to the sorts of error vectors just canvassed. Can these potential sources of error be detected, avoided, preempted, mitigated, or compensated for by means of such resources? If so, then it is a fair point, and my arguments would need to be reconsidered with these other resources counted in CAMC's favor. But if not, then these further resources can do nothing to blunt my argument for the necessity of x-phi.

3. *A defense of modest x-phi methodological norms for philosophy*

So far, I have argued for what benefits that x-phi can at least potentially bring, of general methodological value: CAMC has deficiencies, in the form of vectors of inappropriate sensitivity, and x-phi can go some distance towards wrangling those vectors under control. It is yet a further step to say that x-phi is methodologically *necessary*, though. Sometimes potential benefits are not worth the expected costs, and any proposed change to our methodological norms would have to be subjected to a calculation of the expected value of that trade-off. For example, sometimes philosophers accidentally commit formal fallacies in their papers – surely rather rarely in this day and age, but still with greater than zero frequency, a philosopher will trip over a scope ambiguity or an unintentional inversion of quantifier

⁸ See Alexander (forthcoming), Buckwalter (forthcoming), and Nado (2014) for overviews. Key recent results include Schulz et al. (2011), Schwitzgebel and Cushman (2015), and Tobia et al. (2013).

order. Were we to adopt a norm requiring the translation all of our arguments rigorously into an appropriate formal language and then doing explicit, axiomatized derivations, then perhaps it would reduce the number of those fallacious arguments, maybe even to zero. Yet it does seem that the expenses incurred in following such a norm would swamp the value of any such decrease in fallacies. (Just think of how much more painful paper refereeing would be!) And there are not just practical costs, but epistemic ones as well: even assuming (as is not terribly implausible) that any errors in formal derivation would be caught prior to publication, nonetheless there would be an increased risk in introducing errors into papers during the step of translating into and out of the formal calculi. We would be deprived of the contributions of philosophers who were otherwise insightful and skilled but lacking in technical chops. All in all, we might well be curtailing one vector of errors by introducing still worse ones. Even the biggest proponents of formal methods in philosophy would, I think, freely endorse the claim that such a “norm of universal derivation” would not be wise for philosophers to adopt.

It is thus entirely appropriate to ask how any proposed change to incorporate λ - ϕ into our methodological norms would fare in such a cost-benefit analysis. How would a “norm of universal experimentation” score, in such terms? Actually, it would fare at least as poorly as its formal counterpart, for all the same reasons *mutatis mutandis*. Such a norm falters upon the uneven distribution of the relevant aptitudes in the profession; the increased risk of error at the stage of operationalization and design of materials when trying to test those claims that are not especially amenable to experimental treatment; even higher practical costs, since it is generally much more expensive to run a good study with adequate power, than it is to work out a formal proof; and so on.

We have done fairly well with more moderate norms for the operation of formal tools in the philosophical workshop, and these well-implemented philosophical tools could thus serve as models for the installation of experimental tools there as well. While we do not require across-the-board formal derivations, we do possess a reasonably good sense about what sorts of inferential steps may be so tricky as to benefit from a more mathematical treatment. For example, a paper claiming a non-obvious entailment from even a small but moderately complicated set of propositions with, say, iterated modal operators, or even just a handful of nested quantifiers of first-order logic, will likely be required to include at least enough machinery of proof to make the entailment perspicuous. And of course we have usefully entrenched norms as to how such proofs are to be presented and notated, and the success of such norms is scaffolded by the inclusion of logic courses in nearly all graduate programs in philosophy today, with substantially

more mathematical training readily available for those who find it relevant to their own projects. In short, we have a good working understanding of where logical tools can be helpful, both where we may be prone to various sorts of errors without its aid, and how to use it to overcome those liabilities; we have conventions for how to report on the operation of those tools in our publications; and we have educational practices and expectations in place to make sure that this understanding and conventions are widespread.

Many x -phi norms more moderate than “universal experimentation” can follow that model. We can draw heavily from scientific psychology, aided where necessary by “negative program” x -phi, to learn where unaided human cognition may be unacceptably susceptible to error when engaging in philosophical argumentation. For a great many such potential foibles, good social sciences methodology will already offer excellent resources for overcoming them. As noted above, while CAMC offers practically no resources for addressing error vectors like order effects, or outlier effects, especially when intensified by motivated cognition, it is nonetheless very easy to control for order in an experimental design, and outliers can similarly be easily detected, so long as samples are gathered well. (Not that this would *totally* get rid of these problems, especially motivated cognition, which has proved an incredibly thorny problem – probably nothing could do so, and indeed the *total* removal of any source of error is probably a Cartesianly unachievable demand. What we are looking to show here is that there could be norms that incorporate experimental philosophy more generally into philosophical practice, where the expected benefits exceed the costs. In this particular spot, even mitigating the threat of motivated cognition would be a significant benefit, even if it falls unfortunately far short of totally eliminating it.)

So the first set of norms to consider that would require experimental methodological interventions would be targeted specifically to conditions where we expect CAMC’s resources to fall short, on analogy with our norms for requiring formal methodological interventions. That analogy breaks down somewhat when we consider that such norms would be particularly demanding when those expectations have been well-confirmed by the relevant sorts of empirical investigations (and such norms would be accordingly relaxed when such investigations disconfirm any such prior expectations of susceptibility to error; see Mortensen and Nagel forthcoming). The particular sorts of vectors to be checked would likely vary with the target concepts, in accord with the state of the art of our knowledge about them; for example, a philosopher looking to retail a free will attribution case would want to check for possible variation according to introversion/extroversion of the attributor, since there are robust results indicating the existence of such an error vector in this domain (Feltz and Cokely

2009), but there might not be a need to do so with knowledge attributions at this time.

Another norm to consider might be summarized, “practice good defensive x -phi”. When a philosopher wants to put significant argumentative weight on a specific verdict about a novel case, especially one that has not yet been empirically investigated at all, perhaps that philosopher should be required to do some very preliminary, and indeed even fairly superficial work, just as a check on the most common sorts of error vectors – even in the absence of any specific, positive expectation of a susceptibility to error at that particular locus. If the costs of doing x -phi comes down (see below), it should not be burdensome to run a few different variations of any such case, considered in different orders against perhaps a standardized set of anchor cases, and checked across a reasonably diverse subject pool. (I would note in particular that this would require nothing at all complicated in terms of statistics.) While such quick self-checks would not be taken as any sort of definitive demonstration that the desired verdict was the verdict about the case, nonetheless the philosopher offering the case could rightly feel increased confidence that the work was shielded from some of the more common sorts of errors.

There are other kinds of norms we will need to consider adopting as well, beyond just those encouraging or mandating the application of experimental tools. Our professional educational and training norms may also need revision. Just as we require all our PhD’s to be conversant in logic but expect only a few to become specialists in it, I suspect that our profession would be best served with a universal minimum plus support for wide range of more advanced levels of training. Some philosophers who decide they need more advanced experimental tools should also have the option of pursuing collaboration with specialists in the social sciences who are already masters of those methods, as is already the case with a great many highly successful philosopher-and-psychologist collaborations in x -phi. There would be at least two important further consequences of such a norm: first, as a larger slice of the profession gains requisite competence in various scientific methodologies, the practical costs of implementing these norms will decrease. The burden being placed on the profession would be distributed across more backs, and would thus be lighter for all bearing it. Second, as we increase both the number of philosophers with any competence in this area, and the average competence of those that do, we will see an accompanying boost in the quality of the experimental work being done, and deeper benches of referees for journals to help keep the quality high, and improving. And thus the methodological costs of such x -phi norms will also go down, should such educational norms be adopted, because we will face an ever-lower risk of new errors being introduced by the experimentalists themselves.

Just as we would expect different levels of expertise to be inculcated across the professional population, so too might we allow these norms to be implemented in accord with a division of intellectual labor. It may be fine for many philosophers to go about their intuitive business without trafficking in x-phi at all, so long as they are in good, responsive contact to other members of the professional community who are utilizing those methods as needed. In cases where no specific worries have yet been raised, we could probably get away with an “innocent until suspected guilty” norm, so long as the bar for suspicion is set fairly low. It should not take much more than sincere dissent about a case verdict in a thought-experiment to send both parties to their respective labs, or to seek out the help of their friends who have them. (Perhaps this norm could be abbreviated as “trust, but verify”.)

I am just sketching some possible norms here in broad outline, and I am sure that there should be others to consider, both along these lines but either more or less demanding, or concerning other aspects of the profession altogether (such as, say, norms of authorship credit). I hope to have made two points clear by this exercise, though. First, it is useful to see that there are a range of methodological norms we might choose to adopt, as a profession, in order to reap the benefits of x-phi but without anything like the counterproductive extravagance of Universal Experimentation. I think much resistance to x-phi originates in a fear that experimentalists must have something just that crazy in mind! It may well be that we could have a nice bit of methodological improvement at what would really be a rather low professional cost. Second, although the norms sketched above are all very modest, they nonetheless remain dangerous to any conception of philosophy as an armchair discipline. In our terms here: adoption of even these modest norms would represent a significant departure from CAMC, conceived at the level of the profession on the whole. Any *particular* armchair-residing philosopher may perhaps be licensed to remain thus seated – but at best, only so long as they are in the right kind of responsive contact with those who are not. Some individual philosophers can pretty much restrict themselves to CAMC, so long as philosophy on the whole does not, and so long as our methodological resources expand in ways that facilitate our detection of and compensation for the sorts of errors that CAMC may be unknowingly prey to.

4. *On the necessity and sufficiency of modest x-phi methodological norms*

Both proponents of x-phi and defenders of the armchair may well wonder, however, whether norms so low in cost as these could still provide enough benefit to warrant our adopting them, let alone necessitate such adoption. A more radical “experimentalist” might object that in order more

fully to root out these sorts of errors, we still need something strong, even if not quite as severe as a norm of Universal Experimentation. They might insist on what we might call a *norm of precertification*: an intuition may only be relied upon if we antecedently have significant positive expectation that it will be immune from any established sorts of error vectors. On the other hand, resolute defenders of the armchair – let’s call them “cathedrists” – might wonder if such modest methodological changes, with accordingly modest benefits, would still be worth the fuss, and at such costs as steering some fraction of our graduate students into the requisite sorts of statistical training, or having to expand the referee pool for major journals to include persons with such training. So I will argue now that even modest norms like those proposed above could yield greater benefits than they may seem *prima facie* to these (perhaps hypothetical) resolutely partisan participants in these debates. But to get to that point, I first need to make some big-picture remarks about the nature of philosophical inquiry (or, rather, about the nature of the particular kind of philosophical inquiry in which these sorts of intuitive methodologies under consideration are deployed).

All methodologies have to engage with the fallibility of any human endeavor, and communities of inquiry have developed two distinct strategies for dealing with the threat of error. One strategy is to impose highly demanding constraints on when a method will count as having successfully delivered a result, such that it will count vanishingly few errors among its deliverances. The most obvious advantage of this kind of strategy is that one can place enormous trust in those results, and in turn, those results can be built upon in further investigations with almost no fear of being thereby led astray. The stockpile of certified results can be expected to grow almost entirely monotonically. But this sort of approach, which I have called “M-methodology” (Weinberg 2015), has at least two major drawbacks as well. First, in order to achieve this state of near-infallibility, the constraints may be so demanding as to be navigable by only a narrow and elite set of investigators. Second, the nature of M-methodologies can preclude delivering any results that it can reckon only as merely probable, even if that probability is rather high. To do so would allow in too great a risk of error, compounding as other probable-but-not-close-enough-to-certain results are also included among its deliverances – soon negating that very advantage that was the methodology’s key selling point.

In domains where we want or even need to use more probabilistic and nonmonotonic forms of inquiry, we instead deploy what I termed “S-methodologies”: we accept that a number of errors may be allowed in at any time, and in contrast with M-methodologies, we must invest significant resources in rooting them out afterwards. Of course significant measures will still be taken to try to keep them from creeping into our findings in the

first place, but we do not merely resign ourselves to the imperfections of such preventative cognitive defenses, and to whatever falsehoods may thus get past them: we take measures to find them and push them back out the door. Resources used for this after-the-fact detection of and correction for errors, I have called *hope*.⁹ Hope is a forward-looking methodological virtue, concerning what we can do to get rid of mistakes we may already have made or will someday make, and it should be understood in contrast with (but not at all contradictory to) more traditional epistemological virtues like process reliability.

We should ask, then, just how much hope any given methodology from the S-family should be required to have, in order to be counted as in good standing. Let me suggest that the normative demand for hope is proportional to at least the following three factors:

- (1) the extent to which one is using an S-methodology not an M-methodology;
- (2) the actual risk of errors in one's evidence set;
- (3) the sensitivity of one's inferences to errors.

Moreover, one's resources for the mitigation of error must be able to address the particular kinds of errors for which one is at risk. (One significant mistake in my earlier treatment of hopefulness was treating the demand for hopefulness without this sort of differentiation in terms of the particular risks of error.)

In short: the more your ultimate theoretical products can be disturbed by errors, the more you need to take active steps to reduce the risk of such disturbance by incorporating effective resources for detecting and mitigating those errors.¹⁰

Turning now to philosophy in particular, my suspicion is that we are an S-methodological field, but that has a more M-methodological self-understanding, at least in this vicinity. We valorize proofs and deductions, and as noted above, training in such tools is one of the few universal *sine qua nons* of PhD programs today. And, just to be very clear, I do not mean at all to be diminishing the value of such tools! And I have no objections whatsoever

⁹ In my (2007); I would note that the larger argument there, against armchair methods, appeals to a principle about hope that I now think is too strong, convinced by arguments such as Brown (2013), Grundmann (2010), and Ichikawa (2012). This section of this paper can be considered a revision and update of that argument.

¹⁰ There is an interesting question here as to what to do should one find oneself facing such a high demand for hope, but with no means available to meet that demand (Brown 2013). We can set that question aside here, however, since this controversy about x-phi vs. armchair turns on whether philosophical methodology must be expanded to include resources that are, in fact, readily available, namely, those of the social sciences.

er to raise to logic or other more mathematical subdisciplines, such as formal epistemology. But considering analytic philosophy more broadly, we allow too many sources of evidence and modes of inference that are too fallible, too probabilistic, to count ourselves as practicing an M-methodology. Even without the x-phi results, we knew our intuitions to be at least modestly fallible; much philosophy draws nontrivially upon the sciences, which are almost definitionally S-methodological; inferential tools like reflective equilibrium and inference to the best explanation proceed nonmonotonically, and often require us to backtrack and revise. Even introspection, with its sometime promise of first-person authority, falls short of M-methodological standards of immunity from error (Schwitzgebel 2008).

Moreover, we have seen already that CAMC does not possess adequate resources on its own to take adequate stock of its overall actual risks of error, and a number of x-phi findings are at least highly suggestive that that risk is very real and as yet unmitigated. Now, perhaps these considerations are already enough by themselves to motivate a high requirement for hope in philosophical methodology. We are a field that is at risk of error, and that risk is higher than our current standard methods have antecedently reckoned, and more, it seems, than they can handle. Yet when we consider the nature of philosophical inference today, we can see that that demand must be higher still. Jennifer Nado discusses the *epistemic demandingness* of different sorts of inquiry, and argues that philosophical inquiry, which so often is framed in terms of exceptionless universals, is enormously demanding. That is, many modes of philosophical inference – including, with respect to this special issue’s topic, philosophical analysis – require a much higher degree of reliability than other modes of cognition, especially many that operate just fine for the purposes of our ordinary lives. She illustrates with the following useful example:

Consider a group of 10 objects, a, b, c ... j, and two properties, F and G. Now consider a subject who possesses a “folk theory” devoted solely to those objects and their properties, on the basis of which the subject makes judgments regarding the applicability of F and G to the objects in the group. Suppose that, by means of this folk theory, our subject produces the judgments Fa, Fb, Fc ... Fj, and the judgments Ga, Gb, Gc ... Gj. Finally, suppose that in actuality, \sim Fa and \sim Gb – all other judgments are correct. Out of 20 judgments, the subject has made 18 correctly – she is, then, a reasonably reliable judge of F-hood and of G-hood on the cases to which her folk theory applies. We would likely say that it is epistemically permissible for the subject to rely on such judgments in normal contexts.

Suppose, however, that our subject is a philosopher; further, suppose her to be concerned with the nature of F-hood and of G-hood. Our subject might then come to hold certain theoretical claims about the nature of F-hood and G-hood on the basis of those initial classificatory judgments. She might, for instance, infer that eve-

rything (in the toy universe of 10 objects) is F, that everything is G, and that if something is F then it is G. She would be wrong on all counts. The example is simple, but it shows that a certain principle – that the general reliability of one’s classificatory judgments directly entails the general success of one’s theory-building – is clearly false. Generating an accurate theory is highly epistemically demanding; an otherwise respectable source of evidence may not suffice. (Nado 2015, pp. 213-214)

Since Nado focuses on the reliability of the initial judgments, I need to transpose her example into the more future-directed key of hope. (This is no criticism of her arguments, but should be seen rather as supplementary to them.) Her hypothetical philosopher of the second paragraph could still actually be in fine shape, if(f!) she has the error-detecting resources to discern that she was wrong about Fa and Gb, and can thus at least down the road come to retract her “theoretical claims about the nature of F-hood and G-hood”. This is still an instance of a very high level of epistemic demandingness, but imposed now upon the degree of hope that the philosopher must require of her practices, given that her reliability in the first place is only 18 out of 20, but her theories require her to be right about all 20. (It would also be a disaster if she erred without correction the other way: if all F’s necessarily are all G’s, but she has an uncorrected error that b is F but not G.)

We thus find ourselves with (i) an S-methodology that (ii) is exposed to significant risks of error for which it lacks the means to address within its consensus set of methodological resources, but which (iii) uses modes of inference that are rather epistemically demanding, that is, which can easily go awry with even a small number of errors among its premises. The normative demand for hope is thus high, but unfulfilled.

We are now ready to return to the radical experimentalist and cathedrist, and their concerns about whether we should adopt anything like the modest x-phi norms sketched in section 3. In response to the radical experimentalist, who demands that any intuition about a case be established as error-free before we can make use of it, we can see now that it would be a mistake to think that such errors need to be stamped out beforehand. That sort of precertification argument is entirely appropriate to M-methodologies, but not at all to S-methodologies. If I am right about hope, then it is perfectly kosher for our resources to correct any missteps in the course of inquiry after the fact, as part of our ongoing investigative journey. And thus, in particular, no philosopher needs to sit around waiting for the x-phi folks to wrap up their studies, before getting on with their work. So I am in agreement with Williamson when he argues that “it is a fallacy to infer [from the some-time philosophical relevance of experiments] that philosophy can nowhere usefully proceed until the experiments are done” (Williamson 2007, p. 6) – so long, however, as that philosophical work will

over time be in good, responsive contact with the kind of experimental work that may help to reveal any errors they may be prey to now without realizing it.

If that is correct as a response to a radical experimentalist's demand for precertification, then note that it also contains the seeds of a reply to the cathedrist as well. Hope may be all that an M-methodology needs with regard to error vectors, and not precertification – but it is a robust requirement nonetheless. And if the above suggestions about the normative demands of hope are correct, then it is a requirement that falls squarely upon the philosophical community, but cannot be discharged using only the resources made available by CAMC. Experimental philosophy will, indeed, be necessary, even if only in the modest form sketched in section 3.

5. Conclusion

In all then, we should expect that the normative demand for hope in philosophical methodology is rather high. Yet this hope cannot be sufficiently supplied by resources internal to CAMC. My contention here is that nonetheless modest x -phi methodological norms could go a long way to providing that hope, were they broadly incorporated into our practices. The epistemic demandingness of many of our modes of inference in philosophy make clear how a little bit of benefit in catching errors may be worth even a fair amount of cost, since that demandingness magnifies the costs of undetected errors in the first place. And I have tried to suggest how the costs of implementing such norms need not be so high as some have feared, as well.

There may be a kind of general methodological principle to explore here: a methodology can only be autonomous if its own risks of error are included in its closure. CAMC has proved illusory as an autonomous set of methodological resources for philosophy, precisely because its deliverances are susceptible to error vectors that lie outside of its own methodological ambit. We need a new methodological consensus, one that includes these needed resources from the sciences that will allow greater error-detection and error-correction.

Department of Philosophy & Program in Cognitive Science
University of Arizona
United States

E-mail: jmweinberg@email.arizona.edu

References

- Alexander, J. forthcoming: "Philosophical Expertise", to appear in J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy*, Oxford, Wiley-Blackwell.
- Brown, J. 2013: "Intuitions, Evidence and Hopefulness", *Synthese*, 190(12), pp. 2021-2046.
- Buckwalter, W. forthcoming: "Intuition Fail: Philosophical Activity and the Limits of Expertise", to appear in *Philosophy and Phenomenological Research*.
- Buckwalter, W., Knobe, J., Nichols, S., Pinillos, N., Robbins, P., Sarkissian, H., Weigel, C. and Weinberg, J. 2012: "Experimental Philosophy", *Oxford Bibliographies Online*, 1, pp. 81-92.
- Cappelen, H. 2012: *Philosophy without Intuitions*, Oxford: Oxford University Press.
- Deutsch, M. 2010: "Intuitions, Counter-examples, and Experimental philosophy", *Review of Philosophy and Psychology*, 1, pp. 447-460.
- Feltz, A. and Cokely, E.T. 2009: "Do Judgments about Freedom and Responsibility Depend on who You Are? Personality Differences in Intuitions about Compatibilism and Incompatibilism", *Consciousness and Cognition*, 18(1), pp. 342-350.
- Fischer, E. and Collins, J. (eds.) 2015: *Experimental Philosophy, Rationalism, and Naturalism: Rethinking Philosophical Method*, London-New York, Routledge.
- Grundmann, T. 2010: "Some Hope for Intuitions: A Reply to Weinberg", *Philosophical Psychology*, 23(4), pp. 481-509.
- Hales, S. 2006: *Relativism and the Foundations of Philosophy*, Cambridge, MA, MIT Press.
- Ichikawa, J. 2010: "Explaining away Intuitions", *Studia Philosophica Estonica*, 2, pp. 94-116.
- Ichikawa, J. 2012: "Experimentalist Pressure against Traditional Methodology", *Philosophical Psychology*, 25(5), pp. 743-765.
- Knobe, J. forthcoming: "Experimental Philosophy is Cognitive Science", to appear in J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy*, Oxford, Wiley-Blackwell.
- Ludwig, K. 2007: "The Epistemology of Thought Experiments: First Person versus Third Person Approaches", *Midwest Studies in Philosophy*, 31, pp. 128-159.
- Machery, E., Mallon, R., Nichols, S. and Stich, S.P. 2004: "Semantics, Cross-cultural Style", *Cognition*, 92(3), pp. B1-B12.
- Mortensen, K. and Nagel, J. forthcoming: "Armchair-friendly Experimental Philosophy", to appear in J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy*, Oxford, Wiley-Blackwell.
- Nado, J. 2014: "Philosophical Expertise", *Philosophy Compass*, 9, pp. 631-641.
- Nado, J. 2015: "Intuition, Philosophical Theorizing, and the Threat of Skepticism", in Fischer and Collins (2015), pp. 204-221.

- Nolan, D. 2015: "The a posteriori Armchair", *Australasian Journal of Philosophy*, 93, pp. 211-231.
- Prinz, J. 2008: "Empirical Philosophy and Experimental Philosophy", in J. Knobe and S. Nichols (eds.), *Experimental Philosophy*, Oxford, Oxford University Press, pp. 189-208.
- Schulz, E., Cokely, E. and Feltz, A. 2011: "Persistent Bias in Expert Judgments about Free Will and Moral Responsibility: A Test of the Expertise Defense", *Consciousness and Cognition*, 20, pp. 1722-1731.
- Schwitzgebel, E. 2008: "The Unreliability of Naive Introspection", *Philosophical Review*, 117(2), pp. 245-273.
- Schwitzgebel, E. and Cushman, F. 2015: "Philosophers' Biased Judgments Persist despite Training, Expertise and Reflection", *Cognition*, 141, pp. 127-137.
- Stich, S. and Weinberg, J. 2001: "Jackson's Empirical Assumptions", *Philosophy and Phenomenological Research*, 62(3), pp. 637-643.
- Swain, S., Alexander, J. and Weinberg, J. 2008: "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp", *Philosophy and Phenomenological Research*, 76(1), pp. 138-155.
- Tobia, K., Buckwalter, W. and Stich, S. 2013: "Moral Intuitions: Are Philosophers Experts?", *Philosophical Psychology*, 26(5), pp. 629-638.
- Weinberg, J. 2007: "How to Challenge Intuitions Empirically without Risking Skepticism", *Midwest Studies in Philosophy*, 31, pp. 318-343.
- Weinberg, J. 2009: "On Doing Better, Experimental-style", *Philosophical Studies*, 145(3), pp. 455-464.
- Weinberg, J. 2014: "Cappelen between Rock and a Hard Place", *Philosophical Studies*, 171, pp. 545-553.
- Weinberg, J. 2015: "Humans as Instruments: Or, The Inevitability of Experimental Philosophy", in Fischer and Collins (2015), pp. 171-187.
- Weinberg, J., Gonnerman, C., Buckner, C. and Alexander, J. 2010: "Are Philosophers Expert Intuiters?", *Philosophical Psychology*, 23, pp. 331-355.
- Williamson, T. 2005: "Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking", *Proceedings of the Aristotelian Society*, 105, pp. 1-23.
- Williamson, T. 2007: *The Philosophy of Philosophy*, Oxford, Blackwell.