

August 1, 2014

Forthcoming in *A Companion to Experimental Philosophy*,
Justin Sytsma and Wesley Buckwalter, eds. (Blackwell)

Armchair-friendly Experimental Philosophy

Kaija Mortensen & Jennifer Nagel

Introduction¹

The relationship between experimental and traditional philosophy is often seen as hostile, and not without reason. Scanning the work of prominent experimentalists, we see that one of them is willing to bet that experimental work will show that “a great deal of what goes on in contemporary philosophy, and a great deal of what has gone on in the past, belongs in the rubbish bin” (Stich 2009, 232), while several others hope their results will “make live and salient the possibility that [intuition-driven armchair philosophers] will find that their practice may in fact be built on an unacceptably shifting foundation” (Swain, Alexander, and Weinberg 2008, 153-4). But if remarks such as these have made it seem that experimental philosophy is by its very nature opposed to traditional philosophy, this impression is misleading. Experimental and traditional philosophy certainly differ in their methods, but to say that methods are different is not necessarily to say that they are incompatible, or that one must be pursued at the expense of the other. Although some early work in experimental philosophy may have led philosophers to believe that experimental methods must be opposed to traditional (or ‘armchair’) methods, much recent work has been aimed at rebutting those early challenges, and at reconciling experimental

¹ Thanks to Wesley Buckwalter, Justin Sytsma, and Jonathan Weinberg for helpful comments on previous versions of this chapter.

and traditional ways of tackling philosophical problems. This chapter examines ways in which experimental methods can complement and even strengthen armchair-style philosophy.

Although we believe that experimental philosophy at its best is friendly to traditional philosophy, we consider it important to understand the source of the impression that this is not the case. Section 1 of this paper traces the impression of hostility back to three serious challenges to intuition-driven armchair philosophy that have been raised by experimental philosophers. Because a better understanding of the nature of armchair philosophy helps to show that these challenges do not ultimately pose a fatal problem, Section 2 takes a closer look at the nature of armchair philosophy in general, and the method of cases in particular. We pay particular attention to the question of how experimental methods might bear on the legitimacy of intuition-driven armchair philosophy. Section 3 reviews recent work in experimental philosophy that has effectively defended armchair philosophy from the three major challenges described in Section 1, and Section 4 investigates ways in which experimental philosophy can enhance and extend the power of traditional philosophical methods.

1. Three experimentalist challenges to the armchair

It may be hard to develop a general characterization of armchair philosophy, but it's easy to find clear examples of it. If anything counts as armchair philosophy, it's Edmund Gettier's classic (1963) paper criticizing the classical analysis of knowledge as justified true belief. Gettier develops two intuitive counterexamples to that analysis, and concludes that justified true belief does not always amount to knowledge; his conclusion has been widely accepted among mainstream epistemologists since. But one might wonder about the status of Gettier's crucial

intuitive judgments about his examples: are they in fact objective judgments about the nature of knowledge, well-founded judgments that could be made by any rational individual? In one of the founding papers of the contemporary experimental philosophy movement, Jonathan Weinberg, Shaun Nichols, and Stephen Stich challenged the assumption that Gettier case intuitions are simply objective reflections of the nature of knowledge. On the basis of literature about cross-cultural differences in reasoning styles (e.g. Nisbett et al. 2001), Weinberg, Nichols and Stich (2001) hypothesized that the intuitions Gettier apparently assumed were universal might vary by culture and socio-economic status. After running empirical studies that appeared to show such variation, they drew harsh conclusions about the legitimacy of Gettier's method:

It may well be that upper-middle-class Westerners who have had a few years of graduate training in analytic philosophy do indeed all have strong, modality-linked intuitions about Gettier cases. But since most of the world's population apparently does not share these intuitions, it is hard to see why we should think that these intuitions tell us anything at all about the modal structure of reality, or about epistemic norms or indeed about anything else of philosophical interest. (Weinberg, Nichols, and Stich 2001, p. 452)

The *Diversity Challenge*, as we shall call it, was not limited to intuitions about Gettier Cases. In later work, Stich makes it clear that he sees diversity as a threat to traditional philosophy much more broadly:

For 2,500 years, philosophers have been relying on appeals to intuition. *But the plausibility of this entire tradition rests on an unsubstantiated, and until recently unacknowledged, empirical hypothesis* – the hypothesis that the philosophical intuitions of people in different cultural groups do not disagree. Those philosophers who rely on intuitions are betting that the hypothesis is true. If they lose their bet, and if I am right that the prospects are very dim indeed for producing a convincing theory of error which explains why a substantial part of the world's population has false intuitions about knowledge, justice, happiness, and the like, then a great deal of what goes on in contemporary philosophy, and a great deal of what has gone on in the past, belongs in the rubbish bin. I think it is too early to say with any assurance who is going to win this bet – though if I were a practitioner of intuition-based philosophy I'd be getting pretty nervous. (Stich 2009, 232, emphasis in original)

Notwithstanding his concession that it is too early to be assured of the demise of armchair philosophy, Stich's declaration, "if I were a practitioner of intuition-based philosophy I'd be getting pretty nervous", conveys an expectation that the results of experimental philosophy will be unfriendly to traditionalists.

Specific concerns about cross-cultural diversity in intuitions were also generated by empirical work on intuitions about reference. One study reported that East Asians and Westerners differ in whether they tend to have causal-historical or descriptivist intuitions about the reference of proper names (Machery et al. 2004). The authors took this finding to signal a pressing need to change traditional philosophical practice, summing up the lesson of their findings as follows:

Our data indicate that philosophers must radically revise their methodology. Since the intuitions philosophers pronounce from their armchairs are likely to be a product of their own culture and their academic training, in order to determine the implicit theories that underlie the use of names across cultures, philosophers need to get out of their armchairs. And this is far from what philosophers have been doing for the last several decades. (Machery et al. 2004, B9)

Cross-cultural differences do not exhaust the Diversity Challenge. Further work in experimental philosophy has aimed to show demographic variation in philosophical intuition along such dimensions as gender—where "large, unexpected and dramatic" differences have been said to occur (Buckwalter and Stich 2014)— personality type (Cokely and Feltz 2009), and age (Colaco et al. 2014).

Meanwhile, the Diversity Challenge is not the only way in which experimental philosophy has seemed hostile to the armchair. Noting that armchair philosophers often make assumptions about how people ordinarily think about knowledge, morality and other matters of

philosophical interest, some experimentalists have claimed that their empirical studies of folk thinking have “challenged familiar assumptions, showing that people do not actually think about these issues in anything like the way philosophers had assumed” (Knobe and Nichols 2008, 3). We’ll label this the *Ignorance of Folk Thinking Challenge*. Perhaps the first experimental philosopher to press this kind of challenge was Arne Naess, in the 1930s.² Naess observed that many philosophers theorizing about truth appealed to ‘the common sense theory of truth’, without offering empirical backing for the suggestion that any particular theory was commonly held. Naess conducted a series of surveys designed to map what subjects with no philosophical training said about truth. He found that participants asked to define truth did not converge on any shared theory, raising a challenge about exactly what philosophers were referring to when they spoke of ‘the common sense theory of truth’ (Naess 1938, Naess and Molland 1938). To defend armchair philosophy from the Ignorance of Folk Thinking Challenge, we’ll need to take a closer look at the extent to which work like Naess’s exposes folk thinking, and at the question of how important it is for armchair philosophers to have an accurate understanding of folk thinking in the first place.

We will also discuss one further experimentalist move that can appear to be hostile to the armchair. We will label this move the *Questionable Evidence Challenge*. Even in the absence of demographic variation, and even if philosophers are well aware of how the folk think, it is argued that philosophically significant intuitions may be inappropriately sensitive to such apparently irrelevant considerations as the order in which cases are presented (Swain, Alexander, and Weinberg 2008), the font in which they are written (Weinberg et al. 2012), and details of the temporal framing of the case (Weigel 2011). Summing up this family of problems, Joshua

² For a discussion of Naess’s challenge in its historical context, see Barnard & Ulatowski (unpublished).

Alexander and Jonathan Weinberg write, “these kinds of intuitional sensitivity are both *unwelcome* and *unexpected*, and the very live empirical hypotheses of their existence create a specific kind of methodological challenge to armchair intuitional practices in philosophy” (2014, 133). What Alexander and Weinberg propose is not the complete elimination of reliance on intuition in philosophical theorizing, but a view they call “restrictionism”, which allows philosophers to use as evidence only those intuitions whose stability and trustworthiness has been established through empirical work (Alexander and Weinberg 2007, 2014). According to restrictionists, even if ordinary intuitions—for example, intuitions about knowledge—are in daily life quite likely to be right, for the delicate purposes of philosophy, intuitions constitute questionable evidence, evidence encumbered with problems that cannot be solved from the armchair. If a philosopher wishes to rely on intuitions as evidence, restrictionists contend, the burden of proof is on the philosopher to demonstrate that the relevant intuitions have been properly vetted empirically.

Each of these three challenges has raised the threat that armchair philosophy would need to be revised or abandoned, given certain empirical results. The first way in which experimental philosophy could be friendly to the armchair would be to defend it by building an empirical counter-attack against these three challenges. We’ll examine work in that vein shortly. But we should emphasize from the outset that although anti-armchair challenges have featured prominently in experimental philosophy’s public image, this work is not in fact representative of the bulk of current work in experimental philosophy. Having conducted a statistical analysis of over 400 experimental philosophy articles published between 2009 and 2013, Joshua Knobe (this volume) reports that just 1.1% of these papers argue for some version of the conclusion that intuitions are unreliable as a source of evidence in philosophical research. According to Knobe,

the vast majority of work in experimental philosophy aims to make some positive contribution to first-order philosophical debates about topics like knowledge, freedom and moral responsibility, or to cast light on our ways of thinking about these topics. The use of new experimental methods to probe these questions does not necessarily entail hostility towards old methods, and indeed we'll see in our final section that much of this work is arguably armchair-friendly. But whether we are concerned with the relationship between new methods and old ones, or whether we are concerned with the challenges experimentalists have raised directly against old methods, it will help to have a clearer picture of the nature of those old methods themselves. Our next section takes up this task.

2. What is armchair philosophy?

In attacking 'armchair philosophy' or 'traditional philosophy', critics such as Stich, Weinberg and Machery have focused in particular on the method of cases, the method of using particular scenarios to elicit intuitions about knowledge, freedom, reference, or other topics of philosophical interest, where these intuitions are reported by a philosopher without systematic empirical study of the intuitions of others. It is certainly appropriate to describe the case method as traditional; it has been used in Western and Eastern philosophy for millennia, and continues to play a large part in current philosophy as well.³ It is also important to recognize that the case method is not the only tool in the armchair philosopher's kit: a great variety of methods have

³ The claim that case intuitions are important is sometimes disputed, for example by Herman Cappelen (2012). Cappelen observes that philosophers typically construct arguments in support of their responses to cases, and concludes that the intuitive responses themselves bear no evidential weight. However, the fact that an author wants to find explicit argumentative support for a view does not establish that the intuitive support presented for that view is unimportant (on this point, see Chalmers 2014). One is in a dialectically stronger position if one can offer both kinds of support, and furthermore, the arguments offered themselves often rest on further premises taken to be justified intuitively.

some claim to count as traditional, from Plato's dialectical method to Descartes's introspective examination of his ideas in the *Meditations*, to Locke's "plain, historical method" of cataloguing his observations about the ways in which knowledge is acquired and words are used. Formal methods, whether the syllogistic logic of the medievals or more contemporary forms of logic, decision theory, and semantics, should certainly count as traditional in virtue of their well-entrenched place in philosophy. Given that armchair philosophers have these other methods at their disposal, the empirical discovery of problems in the intuition-driven case method does not automatically demonstrate that armchair philosophy is in peril; it may well be the case that problems such as intuitional instability can be solved by the application of other armchair methods such as engaging in dialectical exchanges with others, and checking the logical consistency of various sets of judgments (see, e.g. Williamson 2007).⁴ Furthermore, traditional methods other than the case method can in some cases simply trump intuitions: traditional philosophers can uphold a counterintuitive theory on grounds of simplicity or elegance, for example.⁵ Partly for this reason, it is a mistake to characterize armchair philosophy as dedicated strictly to conceptual analysis: armchair philosophers can be concerned with the nature of knowledge (or freedom, or whatever target is at issue) itself, rather than just 'our intuitive concept of knowledge'. Armchair philosophers are not obliged to construct theories that will capture or accommodate all our intuitions: they can take intuitions elicited by the case method as a defeasible source of evidence concerning the ultimate targets of their inquiry. However, cases matter enough to the armchair that if there is empirical work that can also help to defend the case

⁴ Indeed armchair methods can often show where there are problems with the case method in the first place: when the case method turns up paradoxes, or judgments that are individually intuitive but jointly inconsistent, for example, the armchair philosopher can already see that something has gone wrong.

⁵ However, it is noteworthy that the best-known defense of this path—Brian Weatherson's (2003) paper on the attractions of the Justified True Belief (JTB) theory of knowledge—the hypothetical possibility of rejecting the Gettier intuition is motivated by a desire to better accommodate a larger body of other intuitions about knowledge.

method from challenges, remedy its faults, or extend its reach, this work would clearly qualify as armchair-friendly.

There is some controversy about what philosophers mean by ‘intuition’, but we take intuitions to be non-perceptual judgments—or inclinations to judge⁶—that are produced without explicit reasoning. There is a long tradition of understanding the intuitive along these lines in philosophy. For example, Locke (1689) contrasts intuitive with demonstrative reason, where demonstrative reason runs through a series of consciously accessible stages, in contrast to the immediacy of the intuitive. Edmund Burke (1790) characterizes the intuitive as proceeding “without any elaborate process of reasoning.” This understanding of ‘intuition’ also fits with the dominant contemporary model of intuition within psychology, the dual process theory (DPT) view of the contrast between intuitive and reflective cognition (e.g. Evans and Stanovich 2013). Reflective judgments are made by reasoning through a series of consciously accessible stages held in working memory, for example when doing a complex arithmetical problem; intuitive judgments are made without the presentation of consciously available contents in sequential reasoning. Although intuitive judgments do not require explicit sequential thinking, they can still integrate a variety of subtle information, as for example in face recognition. In intuitive judgments about philosophical cases, such subtle information might include the perceived evidential positions, interests and perspectives of the agents we are called upon to evaluate.

The intuitive judgments at issue in the case method are categorization judgments: is this an instance of knowledge? Is this act morally acceptable? Categorizations are not always intuitive: they can be performed reflectively, when we hold a template for the category itself in

⁶ For simplicity, we refer to intuitive judgments in what follows, but we grant that in some contexts intuitive mechanisms may produce something less than a settled judgment, for example an inclination to judge in a way which one reflectively rejects. Nothing in our argument depends on the distinction here.

working memory. For example, if it is explained that a misdemeanor is defined as any crime that carries a maximum sentence of one year or less, and then noted that the crime of reckless endangerment of property carries a maximum sentence of 180 days, one's subsequent categorization of reckless endangerment of property as a misdemeanor will (presumably) be reflective. This proposition is not necessarily judged reflectively by everyone: it could be judged intuitively by a paralegal familiar enough with the defining characteristics of the category not to need to call to mind any explicit definition of a misdemeanor. When it is not performed on the basis of a consciously available template for a category, categorization is intuitive. We still hold relevant features of the case in mind, in working memory, as input to our introspectively inaccessible processing, but the crucial processing stage is intuitive: we do not perform our categorization itself by matching those features to a consciously presented template for the category. One of the main reasons why philosophers employ the case method is as part of the search for a consciously available template, or elements in such a template, or as part of the effort to rule out unsatisfactory candidates for that template: as philosophers, we can raise the question of what it is to be the same person, or to have knowledge, and search for some explicit answer, or components of an answer. If we already had an evidently acceptable definition available to consciousness, we would have much less use for the case method.

The case method assumes that intuitive judgments philosophers invoke about a topic of interest—knowledge, freedom, or whatever is at issue—will tend to reflect their targets. Justin Sytsma and Jonathan Livengood have observed that this method presupposes at least that intuitive judgments are suitably uniform across some relevant group, a presupposition they label “the uniformity conjecture” (Sytsma and Livengood 2011). When the relevant group has been identified it is, as they point out, an empirical question whether the uniformity conjecture is true

for a given class of intuitions. Traditional philosophers are not all agreed on the composition of the relevant group, however, and it is possible that for different philosophical questions, different groups will be relevant. In some domains, for example knowledge attribution, it is possible that the philosopher is tapping a capacity shared by all rational adults: knowledge attribution is, after all, a phenomenon present across all human cultures and heavily used in daily life. If this is right, then the traditional philosopher who consults her own epistemic intuitions is arguably doing something very similar in kind to the experimentalist, essentially running a small experiment (e.g. Nagel 2012). If the same underlying human capacity produces the judgments of philosophers and laypeople alike, it is still possible that philosophers show stronger convergence in their judgments thanks to differences in performance, as opposed to competence: philosophers with an interest in a case are more likely to read the stipulations closely and construe the scenario in a way that makes sense in the dialectical context of the argument. But on the ‘universal capacity’ view of intuitions, we should expect broad similarities in the ways in which philosophers and laypeople respond to cases: the signal is the same—and is arguably a reflection of the philosopher’s target—even if it is mixed with more noise in lay responses. Not all philosophical topics lend themselves to natural human capacities with equal facility, however. Knowledge attribution is a common feature of our everyday thinking, but perhaps judgments involving fine-grained mereological composition principles may be made only by those with appropriate theoretical training. Whether because of differences in competence, or because the judgments in question require special training, defenders of armchair philosophy have allowed that there may be cases for which the traditional philosopher’s thinking will tend to differ from that of the untrained population (e.g. Pinillos et al. 2011). Still others seem open to the idea that, for some cases but not others, philosophers’ responses can be known in advance of structured

empirical investigation to align with those of the folk (e.g. Neta 2012). In what follows, we will not attempt to settle what the relevant uniform group is for every possible philosophical question—we are open to the idea that different groups will be relevant for different philosophical questions—but we will observe that the strongest kind of uniformity, uniformity across philosophers and laypeople of various demographic groups, does seem to hold for many questions of interest. However, we should keep in mind that even uniformity across a restricted group—for example, trained logicians—can be a meaningful sign of accuracy.

3. Rebutting the three experimentalist challenges, experimentally

3.1 The Diversity Challenge

Over the past fifteen years, experimentalist critics of armchair philosophy have claimed to find dramatic diversity in intuition along a variety of demographic dimensions, including age, gender, personality type and ethnicity. However, to the best of our knowledge, none of these claims has clearly withstood subsequent empirical testing. In the first major statement of the Diversity Challenge, Weinberg, Nichols and Stich (2001) claimed to find systematic variation by culture and socio-economic status in responses to epistemological scenarios. For Gettier cases in particular, a series of subsequent studies have failed to show statistically significant cross-cultural differences (Nagel, San Juan, and Mar 2013, Turri 2013, Kim and Yuan unpublished). A more thorough replication study examines not only Gettier cases, but also the other epistemological scenarios tested by Weinberg, Nichols and Stich, including the TrueTemp cases, and does so using a variety of in-class and online methods: here, also, no significant cross-cultural differences were found for any case (Seyedsayamdost forthcoming). As far as we know,

the results on socio-economic status have not been replicated either, and there are methodological concerns about the differences in methods Weinberg, Nichols and Stich originally used in polling the higher and lower-status individuals, differences that may explain why the lower-status individuals answered in patterns closer to randomness on harder cases. Meanwhile, empirical work has also confirmed the stability of intuition in the moral domain, with one recent large-scale study showing relatively little variation by gender, politics, religion and level of education (Banerjee, Huebner, and Hauser 2010).

Cross-cultural differences have been reported in responses to questions about reference: most prominently by Edouard Machery and colleagues, who report differing intuitions among Chinese as compared to Western participants (Machery et al. 2004). Their original report shows the results of two types of scenario that figure in the argument of Saul Kripke's *Naming and Necessity* (1972), Gödel-type vignettes and Jonah-type vignettes, with participants judging two examples of each, in English. For the Jonah-type vignettes there were no significant differences in the responses of the two ethnic groups. For the Gödel-type vignettes, Chinese participants were substantially more likely to give responses that Machery and colleagues took to be indicative of adherence to a descriptivist theory of reference, where Westerners were more likely to respond in line with Kripke's own favoured causal theory. The design and interpretation of this research has been criticized extensively: for example, Barry Lam found no East-West differences when Cantonese-speaking participants were given similar but less ambiguous cases in their native language (Lam 2010). Meanwhile, others have observed that the questions Machery asks invite reports of a theory of reference, rather than its use, and that they are problematically ambiguous (Martí 2009, Sytsma and Livengood 2011). Future results may show

otherwise, but at this time there is no solid evidence that the Diversity Challenge presents an obstacle to traditional philosophical methods in the investigation of reference.

On the question of gender, the Diversity Challenge has been pressed most vociferously by Wesley Buckwalter and Stephen Stich, who summarize their findings by saying “the facts we report about gender differences in philosophy are both important and disturbing” (Buckwalter and Stich 2014, 307). Buckwalter and Stich start with a discussion of an unpublished conference presentation (Starmans and Friedman 2009) which they describe as having found “dramatic differences” in the responses of undergraduate men and women to a Gettier case. Concerned, they specifically solicited reports from experimentalists concerning any findings involving gender differences, and conducted a series of studies on Mechanical Turk, of which they report four as showing significant differences correlated with gender. The authors of the conference presentation no longer maintain that there is good evidence of gender differences in Gettier case responses; they themselves have been unable to replicate their earlier results, and in their subsequent published work they have reported no variation correlated with gender (Starmans and Friedman 2012). Other researchers have also failed to find gender differences in responses to Gettier cases (Wright 2010, Nagel, San Juan, and Mar 2013, Seyedsayamdost 2014). The practice of soliciting reports concerning statistically significant gender differences is not a balanced way of finding genuine gender differences; given the threshold for statistical significance in psychology, such differences should emerge in 5% of all studies even if there is no systematic correlation between gender and philosophical intuition, so the fact that some studies turned up differences is meaningless without information on the size of the pool of studies from which they were drawn (which was not supplied). There is some evidence that these reported cases are indeed random rather than robust effects: a direct replication of three of

them—the cases concerning Compatibilism, Dualism and Physicalism—found no statistically significant differences between the responses of men and women (Seyedsayamdost 2014). The gender difference theory then rests heavily on the four cases that Buckwalter and Stich themselves investigated: scenarios concerning a Brain in a Vat, Twin Earth, Searle’s Chinese Room and the Plank of Carneades. Here Hamid Seyedsayamdost (2014) conducted direct replications of these studies, with two separate data sets, but failed to find a significant gender difference for any case, despite generally greater statistical power; he summarizes his results as yielding “strong evidence that women and men do not differ significantly in their intuitions on the cases examined in this study” (2014, 27).

Another factor that has been used to press the Diversity Challenge is age: in a recent study concerning Fake Barn type Gettier cases, David Colaço and colleagues report that “older people are substantially less inclined” to attribute knowledge to the protagonists in these cases (Colaco et al. 2014). Subsequently, John Turri and Joshua Knobe each tried to replicate this effect with larger data sets, but were not able to find any significant difference correlated with age (Colaco 2014).

Personality type is one last focus for the Diversity Challenge. On the basis of a study involving 58 undergraduates, Adam Feltz and Edward Cokely reported a correlation between extraversion and compatibilism (Feltz and Cokely 2009). However, in a substantially larger study involving data from over 800 participants, and a wider range of vignettes with varied content, Thomas Nadelhoffer and colleagues report only a “small but significant” correlation between extraversion and answers to three particular probes relevant to compatibilism, and did not find a correlation between extraversion and compatibilist tendencies more generally (Nadelhoffer, Kvaran, and Nahmias 2009). Nadelhoffer and colleagues were particularly

concerned that Feltz and Cokely were advancing claims about the interaction between personality type and philosophical view based on a single high-affect vignette. This concern could be pressed more broadly against those who have mounted the Diversity Challenge against armchair philosophy: the most dramatic apparent differences often involve a single batch of responses to a single vignette, and claims about systematic variation should ideally be supported with data covering multiple vignettes with varied content. Where we have no clear pre-theoretical reason to expect diversity—as in the cases of age and gender, neither of which are known to interact with knowledge attribution in the non-clinical adult population—researchers would be well advised to proceed with all due caution.

It is possible that various pockets of genuinely problematic intuitional diversity remain to be discovered, but to date the evidence offered by advocates of the Diversity Challenge has not proven to be robust.

3.2 Ignorance of Folk Thinking

According to Jonathan Livengood and Edouard Machery, “metaphysicians often assume without evidence that they know what the folk think, and these assumptions are sometimes wrong in important ways. In the absence of experimental evidence about folk intuitions, metaphysical speculation is *nominally* constrained by so-called ordinary notions of identity, free will, and the like, while *really* it is checked only by the often peculiar intuitions of metaphysicians themselves” (2007, 1). Livengood and Machery have some advice for armchair philosophers: “The folk probably don’t think what you think they think; so rather than guess from the comfort of your armchair, you ought to go out and check” (2007, 126). Many other experimentalists have

pressed similar concerns about the need for systematic empirical investigation, claiming that their findings will be “unexpected and surprising” to armchair philosophers.

The question of how good armchair philosophers are at predicting lay responses is an empirical question, however, and it is a question that some armchair-friendly experimental philosophers have researched. In a study involving 200 professional philosophers, Billy Dunaway and colleagues worked with materials from four prominent experimental philosophy papers whose results had been characterized as surprising (Dunaway, Edmonds, and Manley 2013). Philosophers were asked to predict folk responses to a series of six probes from these papers, and were also asked to opt out of any question whose content they found familiar. In every case, a large majority of philosophers correctly predicted the folk response (accuracy rates for individual questions ranged from a low of 77.3% to a high of 95.8%). Philosophers should not be assumed without argument to be poor judges of folk responses.⁷

There are other ways of responding to the Ignorance of Folk Thinking Challenge. The first advocate of the challenge was Arne Naess, who suggested that philosophers were going wrong in assuming that there was such a thing as ‘the common sense theory of truth’, given that laypeople asked to define truth failed to converge on any given definition. The most straightforward way to reply to Naess’s version of the challenge would be to observe that laypeople might possess a common sense theory without being able to articulate it explicitly. Just as speakers of English have implicit knowledge of grammatical principles even if they are unable to report them accurately, so also laypeople using the word ‘true’—one of the rare words

⁷ As further evidence of the predictability of folk judgments from the armchair, Joseph Ulatowski (unpublished) notes that in a rarely cited and often overlooked lecture, “Good and Bad Human Action,” G.E.M. Anscombe (2006) seems to anticipate the results of the Knobe effect on the judgments of ordinary people.

said to have a precise equivalent in every natural language (Wierzbicka 1996)—might be guided by a common implicit theory in their application of that word to particular cases. The test of common sense would involve distinguishing appropriate and inappropriate uses of ‘true’, just as we distinguish acceptable and unacceptable sentences in our native language. If philosophers use the same implicit theory as commoners do, they could have access to that ‘common sense’ theory through the contemplation of examples.

Dunaway and colleagues also note the analogy between philosophical intuition and implicit knowledge of natural language, drawing attention to the recent work of Jon Sprouse and colleagues. In a parallel to the Ignorance of Folk Thinking Challenge, some critics of ‘armchair linguistics’ had raised concerns that minimal pairs of contrasting sentences are typically evaluated by a single linguist writing an article (or by the small group of peers and editors who check it) rather than by empirical testing of a larger population (e.g. Gibson and Fedorenko 2010). The danger, it was suggested, is that linguists’ intuitions about questions of syntax might be atypical, or biased in the direction of their own theories, and linguists might for that reason be poor reporters of the judgments of laypeople. However, it is an empirical question how well linguists’ syntactic intuitions match those of the general population, and in recent research comparing a very large range of linguist and lay judgments, it seems that the empirical challenge to ‘armchair linguistics’ is hard to sustain, at least in the case of syntax: linguists’ armchair syntactical judgments are an excellent match to those of the general population, replicating under formal empirical testing at rates of at least 95-98% (Sprouse and Almeida 2012, Sprouse, Schütze, and Almeida 2013). Meanwhile, those who have advocated that linguistics keep to a diet of pure formal empirical testing have yet to come up with any point on which syntactic theory would differ if it made this methodological change.

A final version of the Ignorance of Folk Thinking Challenge would address not the products of folk thinking—the predictable or unpredictable judgments made by the folk—but instead the mechanisms behind these judgments. Do we really know what is driving folk judgments (or philosophers’ judgments) about knowledge, freedom and other targets of philosophical inquiry? In our view, this question about underlying mechanisms is not inherently a hostile question for armchair philosophy; in Section 4 of this paper we will examine some ways in which armchair philosophers can benefit from improved understanding of the mechanisms behind our judgments.

3.3 The Questionable Evidence Challenge

Where early advocates of the Diversity Challenge had proposed banishing reliance on intuitions, current advocates of the more subtle Questionable Evidence Challenge propose selective reliance: “the proper evidential role for philosophical intuitions is one that can only be viewed clearly from outside the armchair, and both bounded by and grounded in a scientific understanding of them” (Alexander and Weinberg 2014, 134). Alexander and Weinberg consider intuitions to be problematically sensitive: “We want our sources of evidence to be sensitive, of course, but we want them to be sensitive to all and only the right kinds of things: that is, whatever is relevant to the truth or falsity of the relevant set of claims” (2014, 132); they then observe that philosophical intuitions are sensitive to “far more than just ethnicity, gender and order effects”; specifically, they mention Feltz and Cokely’s findings on personality (Feltz and Cokely 2009), alongside work on temporal framing and font. Wanting sources of evidence to be sensitive *only* to the truth or falsity of what is relevant is wanting something very strong indeed;

as Alexander and Weinberg rapidly acknowledge, this is not a condition met by perceptual evidence for example. The key difference, in their opinion, is that “we have a pretty good understanding of when sense perception goes wrong” (2014, 133). For philosophical intuitions, on the other hand, they contend that we experience “kinds of intuitional sensitivity that are both *unwelcome* and *unexpected*” (2014, 133), citing Colaço and colleague’s work on age as an example.

Armchair-friendly empirical work here could show that Alexander and Weinberg are overestimating the seriousness of the variation here, or that they are underestimating the capacity of armchair philosophers to spot problems from the armchair. On the first frontier, we have already discussed empirical work showing that Colaço’s findings do not replicate, and that Feltz and Cokely were premature in claiming a clear correlation between extraversion and compatibilism. On the second, there is empirical work suggesting that we do have considerable armchair access to the line between solid and problematic intuitions; in particular, there is relevant psychological work on the relationship between confidence and consensuality in intuitive judgment (where by ‘consensuality’ we mean the extent to which an intuition is shared in the general population). Asher Koriat and colleagues have argued that across a range of perceptual and intuitive forms of judgment, high individual confidence in a judgment correlates well with high stability in that judgment across time and across persons; lower confidence on average is felt for minority judgments, and for judgments that are unstable (Koriat 2011, Koriat 2012, Koriat and Adiv 2011). Studies showing that confidence serves as an effective armchair guide to intuitional stability have been conducted by Jennifer Cole Wright (2010), focusing specifically on intuitions about thought experiments in epistemology. In epistemological practice, judgments that are more widely shared—like judgments about false lemma Gettier

cases—have a stronger dialectical status than judgments that are more contested, like judgments about TrueTemp, or even Fake Barn cases. If a judgment is delicate enough that a change of font will reverse it, it is unlikely to bear much weight in a heated philosophical debate.

There are other ways of responding to the Questionable Evidence Challenge. One way would be to argue that philosophers' answers are on a firmer foundation than those of laypeople, perhaps because they have thought harder about the questions (e.g. Sosa 2007). In support of this hypothesis, Angel Pinillos and colleagues have conducted research showing that individuals who are highly reflective and better informed are less likely to generate the problematic pattern of intuitions characteristic of the Knobe effect (Pinillos et al. 2011). Coupled with the finding of Dunaway and colleagues that 83% philosophers correctly predicted that laypeople would produce the characteristic judgments, this experimental work does something to undercut the Questionable Evidence Challenge. At least in some cases, armchair philosophers are better than one might have thought at keeping unwelcome effects at bay, and at expecting what was said to be unexpected.⁸

4. Empirically extending and enhancing the reach of the armchair

While some experimental philosophy vindicates the armchair by revealing that the intuitions of philosophers 'had it right all along', many experimental philosophy projects enhance and extend the reach of armchair philosophizing in other ways.

⁸ A more radical way to answer the Problematic Foundation Challenge would be to question the assumption that philosophical intuition must always be subject to empirical vetting rather than the other way around: perhaps in some instances what constitutes proper empirical method should be determined by philosophical judgments (for example, about the nature of probability, or experience); this line of response lies beyond the scope of the present article, but see Friedman (1997) for some arguments in favor of such an approach.

One hallmark of the use of the case method in traditional philosophy is the development and exchange of multiple slightly varied cases in order to get a clearer, more refined picture of the boundaries of the target of inquiry. In this exchange, advocates of different theories try to account for the intuitions that others feel by explaining them away as artifacts (of pragmatics rather than semantics, for example) or by incorporating them into their theories. This same kind of exchange is practiced by many experimental philosophers as they systematically vary minute details of cases to investigate how intuitions about the cases change accordingly. This gives investigators further insight into which elements of each case are eliciting what intuitive judgments. This in turn helps investigators better theorize about the cognitive mechanisms giving rise to intuitive judgments.

For example, compatibilists and incompatibilists about free will often appeal to the folk view of free will to establish which side of the debate has the burden of proof. If the folk are naturally incompatibilists, then it is the compatibilists who must offer stronger arguments for the virtues of their theory, to overcome the downside of being unintuitive to the folk. Eddy Nahmias and colleagues (2006) put the intuitiveness of incompatibilism to the empirical test and found, in their early work, that, contrary to assumptions made from the armchair, the folk seem to be compatibilists. One might think that this experimental project was critical of traditional philosophizing, because it showed that a common assumption made from the armchair was wrong. However, as is often the case in all philosophizing, those holding the criticized view sought to defend their preferred theory against the attack. However, they chose to do so by meeting the critics on their own ground – by conducting further empirical research. Nichols and Knobe (2007) ran another study that suggested that the incompatibilism of the folk was an illusion created by a performance error produced by using concrete scenarios rather than abstract

scenarios. When given survey prompts that eliminated the misunderstanding, the majority of participants corroborated the intuitiveness of incompatibilism predicted from the armchair. However, Murray and Nahmias (2014) report the results of two new studies supporting the original conclusion that the folk are compatibilists. These studies show that incompatibilist intuitions are generated when the participants do not understand ‘determinism’ in the same way as philosophers do. When the survey prompts are improved to ensure that the subjects are understanding the right concept, participants again judge free will to be compatible with determinism.

These examples illustrate a powerful way in which empirical work can extend and enhance the reach of traditional philosophical theorizing. Without the aid of empirical investigation, one can only use one’s own intuitions as a guide to the folk view and make some informed guesses about the influences that may be leading those intuitions in the right or wrong direction. However, experimental methods have the added advantage of being able to systematically vary one variable at a time to try to tease out specific influences at work on our intuitions. This allows experimental philosophers to gain insight into the sources of our intuition by testing their educated guesses and, when surprising results are discovered, to explore new theoretical possibilities that might not have been otherwise considered. In these studies, empirical investigation into the sources of the intuitive judgments being elicited enabled the investigators to better assess whether the conclusions drawn from those intuitions were warranted.

Additional examples of this approach can be found in experimental philosophy of mind. Buckwalter and Phelan (2014) aim to challenge ‘the embodiment hypothesis’, the view that folk psychological judgments attribute mental states to things with unified bodies and not to things

that lack unified bodies. In their first study, Buckwalter and Phelan found that embodiment played little to no role in cuing phenomenal state ascription. Instead, functional information had a large impact on phenomenal state attribution. Buckwalter and Phelan then conducted four additional studies to respond to alternative interpretations of their data. With each study, their understanding of the factors influencing phenomenal state attribution deepened. They were able to replicate their original results, even as they refined their experiments to get rid of possible confounds, giving them a more robust basis on which to ground their theories of mental state attribution than assumptions and misunderstood intuitions.

Reuter, Phillips, and Sytsma (2014) engage in a similar experimental process as they investigate the pain paradox: it is common to think that awareness of pain is perceptual (pains are perceived in specific parts of our bodies and, therefore, we can be wrong about our pains) while at the same time it is common to think that awareness of pain is introspective (pains are perceived only by the one in pain and one cannot be wrong about one's pains). Through a series of surveys designed to understand common judgments about pain from many different angles, Reuter, Phillips, and Sytsma found that the introspective view of pain was far less prevalent among non-philosophers than philosophers of mind make it seem in their writings. In addition, they found only miniscule differences between attitudes towards perception and attitudes towards pain, giving us reason to think that common views of pain are not as paradoxical as philosophers would have us believe. When faced with a surprising result, the experimenters designed new studies to tease apart possible interpretations of the data, refining their understanding of the signal they were getting from the folk. Careful experimental work gave investigators resources by which to better understand the intuitions they were probing, allowing

them to tease out possible confounds in their data, giving them a clearer intuitive signal on which to base their theories.

Alexander and Weinberg (2007) identify two types of projects that are concerned with the sources of our intuitions and the resultant warrant they provide for our philosophical beliefs. They dub the approach exemplified above as ‘the proper foundation view’, which they carefully distinguish from ‘restrictionism’. Both projects are concerned with basing our philosophical inquiries on the best data available, and both of these projects can appear to be somewhat hostile to the armchair in that they empirically test assumptions philosophers make from the armchair, and sometimes these tests end up casting doubt on foundations traditional philosophers take for granted. However, in practice, both proper foundationalists and restrictionists enhance rather than undermine armchair theorizing through the incorporation of empirical investigation.

According to Alexander and Weinberg (2007), those seeking proper foundations note statements made by philosophers from the armchair about what ordinary people think about free will, intentional action, pain, consciousness, etc. They note that these statements play evidential roles in the theories developed in the armchair, so they set out to empirically test what the folk actually think in order to provide a “proper foundation” for future philosophical theorizing. Such empirical work supports the armchair by providing a clearer set of intuitions as inputs for that armchair theorizing. The work of Nahmias and colleagues on free will and the work of Buckwalter and colleagues on philosophy of mind provide important examples of this search for firm foundations.

Even the restrictionist program can be seen as in some sense armchair-friendly: “If we are going to learn what intuitional evidence can be used and when intuitional evidence can be used,

we need to know more about where epistemic intuitions come from, what mechanisms are responsible for producing them, and what factors influence them, and this will require looking to the relevant psychology, cognitive science, and empirically informed philosophy of mind” (Alexander and Weinberg 2014, 133). In other words, experimental philosophy lays the groundwork for any future intuition-driven armchair philosophizing. The goal of restrictionism is, at the end of the day, not to commit armchair theorizing to flames but to advocate the necessity of empirical work if any evidential role for intuitions is to be salvaged from the fact that our intuitions are problematically sensitive to irrelevant factors. While we are hesitant to commit to the strong restrictionist claim of the *necessity* of empirical work to intuition-driven methodologies, we endorse its potential *helpfulness* to traditional methodologies; empirically identifying the noise in our intuitions can certainly help us use them more effectively in our philosophical theorizing, presenting us with additional evidence that we need to account for in our theory or explain away as an artifact.⁹

Some work inspired by the restrictionist impulse to locate distorting noise in our intuitions seeks to describe the influences on our intuitions and posit conciliatory explanations of revealed patterns. These explanations challenge past philosophical theories, but also provide concrete evidence in support of alternative theories, though these alternative theories may be much more complex and/or diverse than we may have predicted from the armchair. For

⁹ It’s not obvious that psychological theories can always tell us whether philosophical intuitions are warranted, but to the extent that they can, this type of work could be armchair-friendly. In some cases, it could be helpful to the traditional philosopher to discover that certain intuitions are defective, as well. If we encounter paradoxes—patterns of judgment that seem individually acceptable but jointly contradictory—it would be good to know if some of these judgments are likely to be the products of known limitations or biases within our intuitive mechanisms. For example, it has been argued that certain problematic patterns of epistemic intuition that have motivated contextualism and skepticism in epistemology could be better explained as the products of the bias of epistemic egocentrism (Nagel 2010; Alexander, Gonnerman and Waterman 2014). If this theory about the mechanisms producing those problematic intuitions proves correct, the armchair theorist of knowledge will have a less complex field of intuitive judgments to explain.

example, Weigel (2011) investigates the ways in which intuitions about compatibilism and incompatibilism are influenced by whether the subject contemplates the case happening “next week” or “next year”. After mapping correlations between abstract and concrete scenarios and compatibilist/incompatibilist intuitions, and seeking to determine which intuitions are noise and which ones represent ‘the true signal’ about free will, she concludes that it is likely that both sets of intuitions arise from an underlying competence; it is likely that we have competencies in dealing with both concrete and abstract scenarios, which produce conflicting intuitions in these cases (here she builds on Sinnott-Armstrong 2008). Thus, her restrictionist search to eliminate noise in the signal drives her to contemplate theories of free will that can account for both sets of intuitions. Weigel’s restrictionist work enhances traditional theorizing by using empirical research on the mechanisms producing intuitive judgments to gesture towards a new theory of free will.

Similarly, Feltz and Cokely (2009) weigh in on the compatibilist/incompatibilist debate detailed above, not by finding the ‘right’ set of folk intuitions, but by questioning the notion that the folk have one unified view at all.¹⁰ They argue that “examining and identifying more of the many sources of variation in folk judgment and intuitions offers the possibility of a greater understanding of the folk as well as a disambiguation of disagreements in some of our classically intractable philosophical debates” (Feltz and Cokely 2009). Thus, by more accurately mapping the landscape of folk intuitions, we can uncover additional resources to enhance the quality of the data available to us as we develop and choose between possible theories.¹¹

¹⁰ Feltz & Cokely (2009) cite additional projects of this type.

¹¹ It is unclear whether restrictionists hold that every philosopher must subject his or her intuitions to this scrutiny before proceeding with inquiry or whether they hold the more moderate view that at least someone at the philosophical table needs to be engaging in empirical investigation. It may be that we arrive at better theories if some philosophers are working from the armchair while others are working empirically.

Without doing any empirical studies of his own, Richard Holton (2010) demonstrates another way of conceptualizing the continuity between experimental philosophy and traditional philosophy. Reflecting on a series of studies and follow-up studies on the Knobe effect, Holton offers an explanation and justification of the effect that takes into consideration all of the survey findings but argues, from the armchair, that the Knobe Effect is not a quirk of human psychology, but rather reveals meaningful facts about intentionality. Holton effectively models a way in which armchair theorizing can accommodate experimental data just as it would accommodate data gathered from thought experiments conducted in the armchair.

Experimental philosophy can also resolve a type of impasse sometimes encountered in armchair theorizing. For example, some metaphysicians struggle to reconcile their reflective, expert views about mereological composition with folk metaphysical views to which they think their theories need to be beholden. David Rose and Jonathan Schaffer (submitted) argue on empirical grounds that folk mereology is teleological. But they don't stop there. They then argue (from the 'armchair') that teleological views are false, therefore giving us reasons to believe that folk mereology need not act as a constraint on our philosophical theories of mereology. This empirical work is armchair-friendly because it furthers armchair inquiry by empirically identifying and removing potential roadblocks to that theorizing.

In the Experimental Philosophy Manifesto (2008) Nichols and Knobe identify three goals of experimental philosophy: (1) determine which intuitions are universally shared and which vary according to culture and other demographic differences; (2) investigate what leads us to have the intuitions we do and whether or not the psychological sources undercut warrant; (3) study patterns in people's intuitions to gain insight into how people think.

While we have focused our discussion on the first two types of experimental philosophy, as noted by Josh Knobe (this volume), the majority of extant experimental philosophy studies fall into the third category. According to Knobe and Nichols (2008), "the aim [of this type of experimental philosophy] is to provide an account of the factors that influence applications of a concept, and in particular, the internal psychological processes that underlie such applications. Progress here is measured not in terms of the precision with which one can characterize the actual patterns of people's intuitions but in terms of the degree to which one can achieve explanatory depth" (p. 5). Some (e.g. Sytma and Machery 2013) describe this project as 'neutral'—as opposed to positive or negative—as it runs parallel to traditional concerns regarding conceptual analysis rather than engaging these traditional debates directly. We leave the expert description of these projects and their value to Knobe (this volume).

Some may wonder why these 'neutral' projects should be classified as 'philosophy' at all. While we are not so interested in policing disciplinary boundaries, Knobe and Nichols (2008) do an excellent job defending the continuity between their projects and the history of philosophy. Some experimental philosophy projects acknowledge inspiration from the history of philosophy, and can be seen as an extension of Marx, Nietzsche, and Feuerbach's work on the sources of our religious beliefs as well as an extension of Hume's project as laid out in the first section of the *Enquiry*: "But may we not hope that philosophy, if cultivated with care and encouraged by the attention of the public, may carry its researches still further and discover, at least in some degree, the secret springs and principles by which the human mind is actuated in its operations?" (Hume 1975, 14).

We would like to briefly defend the importance of 'neutral' experimental philosophy projects on different grounds.

One of the great advantages of investigating the mechanisms that give rise to particular intuitive judgments for their own sake rather than for a pre-determined philosophical purpose is that descriptions of the mechanisms can be worked out in isolation from the theories those descriptions might be evidence for or against. The resultant theories of the mechanisms can then be thrown into the ring of theoretical possibilities when philosophers try to understand not just the mind, but also knowledge and other targets of inquiry, offering them more resources for their theorizing than they could formulate if they were univocally focused on their core targets of inquiry rather than also on the way the mind works. It is our view that once intuition data is systematized into various theories, philosophers are then faced with the question of which theory best fits the data. In these cases, work on the mechanisms behind these intuitions that has been proceeding separately from philosophical concerns can provide additional theoretical options, pulling the lab bench into the circle of armchairs without the results of the studies being unduly influenced by philosophical theorizing.

5. Conclusion

We hope to have illustrated a range of possible relationships between traditional armchair theorizing in philosophy and empirical work investigating the ways we think about philosophical topics. For some, these are two autonomous realms, one called philosophy and one called cognitive science. However, we hope to have shown that experimental philosophy in its many forms offers a model of inquiry that enhances what is possible on both sides of the equation. When we put our philosophical intuitions to the empirical test, we will sometimes discover that we are justified in building our theories around our intuitions. Sometimes, however, we will discover that we have misunderstood our thinking about some philosophical topic or other.

Careful experimental work will allow us to develop a better understanding of the mechanisms driving our judgments. This understanding will, in turn, help us develop more accurate, powerful, comprehensive philosophical theories in combination with extant, armchair modes of theorizing.

In the pursuit of these varying projects, experimental philosophers are perhaps divided more by temperament, emphasis, and division of labor than by deep methodological divisions. Some experimental work warns its readers about possible troubles afflicting their use of intuitions in certain domains. Other experimental work emphasizes domains in which intuition use turns out to be unproblematic. Still other experimental work aims to make troublesome intuitions less troublesome by restricting the conditions under which they are used as evidence. All of these projects can be usefully pursued as part of a general shared project of determining where the intuitional signal does, or does not, fall too much prey to noise.¹²

The productive contributions made to first-order debates by experimental philosophers over the past decade give us reason to believe that the constructive value of experimental philosophy is greater than its value as a movement critical of traditional theorizing. However, the critical origins of the movement remind us that progress in philosophy is as dependent on the judgments we weed out as irrelevant as it is on the judgments we endorse as good evidence. We must also remind ourselves that that experimental philosophy proceeds most productively when it sees itself as a move in an ongoing conversation with both the armchair tradition and other empirical studies rather than as an isolated and self-contained method. In addition to engaging in the same kind of case-based dialectic as traditional philosophy, experimental philosophy can be friendly to the armchair when it seeks to deepen our understanding of the cognitive mechanisms

¹² Thanks to Jonathan Weinberg for discussion on this point.

driving our intuitive judgments. These insights can help us assess the evidential value of our intuitions and better understand the nature of philosophical thought.

References:

- Alexander, Joshua, Chad Gonnerman, and John Waterman. 2014. "Salience and Epistemic Egocentrism: An Empirical Study." In *Advances in Experimental Epistemology*, edited by James R. Beebe, 97-117. New York: Continuum.
- Alexander, Joshua, and Jonathan Weinberg. 2007. "Analytic epistemology and experimental philosophy." *Philosophy Compass*, 2: 56-80.
- Alexander, Joshua, and Jonathan Weinberg. 2014. "The "Unreliability" of Epistemic Intuitions." In *Current Controversies in Experimental Philosophy*, edited by Edouard Machery and Elizabeth O'Neill, 128-45. New York: Routledge.
- Anscombe, Gertrude Elizabeth. 2006. "Good and Bad Human Action." In *Human Life, Action and Ethics*, edited by Mary Geach and Luke Gormally, 195-206. Exeter: Imprint.
- Banerjee, Konika, Bryce Huebner, and Marc Hauser. 2010. "Intuitive Moral Judgments are Robust across Variation in Gender, Education, Politics and Religion: A Large-Scale Web-Based Study." *Journal of Cognition and Culture*, 10, 3: 253-281.
- Barnard, Robert, and Joseph Ulatowski. unpublished. Tarski's material adequacy condition and truth as "commonsense".
- Buckwalter, Wesley, and Mark Phelan. 2014. "Phenomenal Consciousness Disembodied." In *Advances in Experimental Philosophy of Mind*, edited by Justin Sytsma, 45-73. New York: Bloomsbury.
- Buckwalter, Wesley, and Stephen Stich. 2014. "Gender and philosophical intuition." In *Experimental Philosophy: Volume 2*, edited by J Knobe and Shaun Nichols, 307-345. New York: Oxford University Press.
- Burke, Edmund. 1790. *Reflections on the Revolution in France*. London: J. Dodsley.
- Cappelen, Herman. 2012. *Philosophy Without Intuitions*. Oxford: Oxford University Press.
- Chalmers, David. 2014. "Intuitions: A Minimal Defense." *Philosophical Studies*.
- Cokely, Edward T, and Adam Feltz. 2009. "Adaptive variation in judgment and philosophical intuition." *Consciousness and Cognition*, 18: 356-358.
- Colaco, David. 2014. More on fake-barn intuitions; replications of Colaco et al. In *Experimental Philosophy*.
- Colaco, David, W. Buckwalter, Stephen Stich, and Edouard Machery. 2014. "Epistemic Intuitions in Fake-barn Thought Experiments." *Episteme*, 11: 199-212.
- Dunaway, Billy, Anna Edmonds, and David Manley. 2013. "The Folk Probably do Think What you Think They Think." *Australasian Journal of Philosophy*, 91: 421-441.
- Evans, Jonathan, and Keith Stanovich. 2013. "Dual-Process Theories of Higher Cognition Advancing the Debate." *Perspectives on Psychological Science*, 8: 223-241.
- Feltz, Adam, and Edward T Cokely. 2009. "Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism." *Consciousness and Cognition*, 18: 342-350.
- Friedman, Michael. 1997. Philosophical naturalism. Paper read at Proceedings and addresses of the American Philosophical Association.
- Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis*, 23: 121-123.
- Gibson, Edward, and Evelina Fedorenko. 2010. "Weak quantitative standards in linguistics research." *Trends in cognitive sciences*, 14: 233-234.
- Holton, Richard. 2010. "Norms and the Knobe effect." *Analysis*, 70: 417-424.
- Hume, David. 1975. *Enquiries*, ed. LA Selby-Bigge and PH Nidditch. Oxford: Clarendon Press.

- Kim, Minsun, and Yuan Yuan. unpublished. "No Cross-Cultural Differences in Gettier Car Case Intuition: A Replication Study of Weinberg et al. 2001."
- Knobe, Joshua. forthcoming. "Experimental Philosophy is Cognitive Science." In *A Companion To Experimental Philosophy*, edited by Justin Sytsma and W. Buckwalter. New York: Blackwell.
- Knobe, Joshua, and Shaun Nichols. 2008. "An Experimental Philosophy Manifesto." In *Experimental Philosophy*, edited by J Knobe and Shaun Nichols, 3-14. New York: Oxford.
- Koriat, A. 2011. "Subjective Confidence in Perceptual Judgments: A Test of the Self-Consistency Model." *Journal of Experimental Psychology: General*, 140: 117-139.
- Koriat, Asher. 2012. "The self-consistency model of subjective confidence." *Psychological Review*, 119: 80-113.
- Koriat, Asher, and Shiri Adiv. 2011. "The construction of attitudinal judgments: Evidence from attitude certainty and response latency." *Social Cognition*, 29: 577-611.
- Kripke, Saul A. 1972. *Naming and Necessity*. Netherlands: Springer.
- Lam, Barry. 2010. "Are Cantonese-speakers really descriptivists? Revisiting cross-cultural semantics." *Cognition*, 115: 320-329.
- Livengood, Jonathan, and Edouard Machery. 2007. "The folk probably don't think what you think they think: Experiments on causation by absence." *Midwest Studies in Philosophy*, 31: 107-127.
- Locke, John. 1689. *An Essay Concerning Human Understanding*. London.
- Machery, Edouard, Ron Mallon, Shaun Nichols, and Stephen P Stich. 2004. "Semantics, cross-cultural style." *Cognition*, 92: B1-B12.
- Martí, Genoveva. 2009. "Against semantic multi-culturalism." *Analysis*, 69: 42-48.
- Murray, Dylan, and Eddy Nahmias. 2014. "Explaining away incompatibilist intuitions." *Philosophy and Phenomenological Research*, 88: 434-67.
- Nadelhoffer, Thomas, Trevor Kvaran, and Eddy Nahmias. 2009. "Temperament and intuition: A commentary on Feltz and Cokely." *Consciousness and cognition*, 18: 351-355.
- Naess, Arne. 1938. "Common-Sense and Truth." *Theoria*, 4: 39-58.
- Naess, Arne, and Einar Molland. 1938. "Truth" as conceived by those who are not professional philosophers. Oslo: Jacob Dybwad.
- Nagel, Jennifer. 2010. "Knowledge ascriptions and the psychological consequences of thinking about error." *Philosophical Quarterly*, 60: 286-306.
- Nagel, Jennifer. 2012. "Intuitions and Experiments: A defense of the case method." *Philosophy and Phenomenological Research*, 85: 495-527.
- Nagel, Jennifer, Valerie San Juan, and Raymond Mar. 2013. "Lay Denial of Knowledge for Justified True Beliefs." *Cognition*, 129: 652-661.
- Nahmias, Eddy, Stephen G Morris, Thomas Nadelhoffer, and Jason Turner. 2006. "Is incompatibilism intuitive?" *Philosophy and Phenomenological Research*, 73: 28-53.
- Neta, Ram. 2012. "Knowing from the armchair that our intuitions are reliable." *The Monist*, 95: 329-351.
- Nichols, Shaun, and Joshua Knobe. 2007. "Moral responsibility and determinism: The cognitive science of folk intuitions." *Nous*, 41: 663-685.
- Nisbett, Richard, Kaiping Peng, Icheol Choi, and Ara Norenzayan. 2001. "Culture and systems of thought: Holistic versus analytic cognition." *Psychological Review*, 108: 291-310.

- Pinillos, N Ángel, Nick Smith, G Shyam Nair, Peter Marchetto, and Cecilea Mun. 2011. "Philosophy's new challenge: Experiments and intentional action." *Mind & Language*, 26: 115-139.
- Reuter, Kevin, Dustin Phillips, and Justin Sytsma. 2014. "Hallucinating Pain." In *Advances in Experimental Philosophy of Mind*, edited by Justin Sytsma, 75-99. New York: Bloomsbury.
- Rose, David, and Jonathan Schaffer. submitted. Folk Mereology is Teleological.
- Seyedsayamdost, Hamid. 2014. "On Gender and Philosophical Intuition: Failure of Replication and Other Negative results." *Review of Philosophy and Psychology*.
- Seyedsayamdost, Hamid. forthcoming. "On Normativity and Epistemic Intuitions: Failure to Detect Differences between Ethnic Groups." *Episteme*.
- Sinnott-Armstrong, W. 2008. "Abstract + Concrete = Paradox." In *Experimental Philosophy*, edited by Joshua Knobe and Shaun Nichols, 209-230. New York: Oxford University Press.
- Sosa, Ernest. 2007. "Experimental philosophy and philosophical intuition." *Philosophical Studies*, 132: 99-107.
- Sprouse, Jon, and Diogo Almeida. 2012. "Assessing the reliability of textbook data in syntax: Adger's Core Syntax." *Journal of Linguistics*, 48: 609-652.
- Sprouse, Jon, Carson T Schütze, and Diogo Almeida. 2013. "A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010." *Lingua*, 134: 219-248.
- Starmans, Christina, and Ori Friedman. 2009. Is Knowledge Subjective? A Sex Difference in Adults. Paper read at 6th Biennial Meeting of the Cognitive Development Society, at San Antonio, Texas.
- Starmans, Christina, and Ori Friedman. 2012. "The folk conception of knowledge." *Cognition*, 2012: 274-283.
- Stich, Stephen. 2009. "Reply to Sosa." In *Stich and his Critics*, edited by D Murphy and M Bishop. Malden, MA: Blackwell.
- Swain, Stacey, Joshua Alexander, and Jonathan Weinberg. 2008. "The instability of philosophical intuitions: Running hot and cold on Truetemp." *Philosophy and Phenomenological Research*, 76: 138-155.
- Sytsma, Justin, and Jonathan Livengood. 2011. "A new perspective concerning experiments on semantic intuitions." *Australasian Journal of Philosophy*, 89: 315-332.
- Sytsma, Justin, and Edouard Machery. 2013. "Experimental Philosophy." In *Encyclopedia of Philosophy and the Social Sciences*, edited by B Kaldis, 318-20. Thousand Oaks, CA: Sage.
- Turri, John. 2013. "A conspicuous art: putting Gettier to the test." *Philosophers' Imprint*, 13: 1-16.
- Ulatowski, Joseph. unpublished. Defending Positive Experimental Philosophy. University of Texas at El Paso.
- Weatherson, Brian. 2003. "What good are counterexamples?" *Philosophical Studies*, 115: 1-31.
- Weigel, Chris. 2011. "Distance, anger, freedom: An account of the role of abstraction in compatibilist and incompatibilist intuitions." *Philosophical Psychology*, 24: 803-823.
- Weinberg, Jonathan M, Joshua Alexander, Chad Gonnerman, and Shane Reuter. 2012. "Restrictionism and Reflection." *The Monist*, 95: 200-222.

- Weinberg, Jonathan, Shaun Nichols, and Stephen Stich. 2001. "Normativity and Epistemic Intuitions." *Philosophical Topics*, 29: 429-60.
- Wierzbicka, Anna. 1996. *Semantics: Primes and universals*: Oxford University Press, USA.
- Williamson, Timothy. 2007. *The Philosophy of Philosophy*. New York: Wiley-Blackwell.
- Wright, Jennifer Cole. 2010. "On intuitional stability: The clear, the strong, and the paradigmatic." *Cognition*, 115: 491-503.