

Experimental Philosophy Within its Proper Bounds

John Symons

Penultimate Version (Forthcoming in *Inquiry*)

11/13/19

Abstract

In *Philosophy Within its Proper Bounds*, Édouard Machery argues that the results of experimental philosophy ('x-phi') should lead us to abandon much of traditional philosophical practice. In its place Machery defends naturalized conceptual analysis as a more modest and pragmatic alternative to standard analytic philosophy. This paper argues that Machery overstates the metaphilosophical significance of x-phi's results. We can and should keep many of the insights and good methodological habits that come with x-phi. However, if one is not already convinced of pragmatism or naturalism, the discoveries of x-phi are unlikely to make too much difference to one's metaphilosophical position.

Introduction

In *Philosophy Within its Proper Bounds*, Édouard Machery argues that the results of experimental philosophy ('x-phi') should lead us to abandon a significant part of

traditional philosophical practice. Specifically, Machery has concluded that the psychological and epistemic condition of human beings is such that we cannot answer philosophically interesting questions about metaphysical necessities. We involve ourselves in intractable disagreements and waste time and energy on problems that fall beyond the proper bounds of human inquiry. His criticism of what he regards as our modal immodesty is informed by 15 years of work in x-phi. He is one of the founding parents of this important sub-discipline and continues to be one its central figures. This important book defends his understanding of the metaphilosophical implications of x-phi.

The change he recommends is significant. If his proposed reform were adopted we would, for example, stop engaging in large parts of metaphysics, epistemology, moral philosophy, and philosophy of mind. Machery recommends reallocating our effort to tasks related to the definitions of concepts and to empirical research into the implications of competing definitions. On his view, this relatively modest enterprise - what he calls 'naturalized conceptual analysis' - is likely to be a more fruitful and a more important kind of inquiry than the work most of us currently pursue.

There is much to praise in the book and some of its criticisms of traditional practice (and the rhetoric surrounding that practice) in analytic philosophy are apt. All things considered, x-phi has made significant contributions to the improvement of philosophical methodology over the past two decades. For an appropriately sympathetic understanding of Machery's metaphilosophical proposals it is helpful to consider them in light of the recent history and sociology of philosophy. I will briefly introduce some of that context before providing a critical assessment of his view.

Metaphilosophy and philosophical methodology are distinct enterprises. As we shall see, the metaphilosophical position espoused by Machery in this book can be distinguished from the contributions of x-phi to philosophical methodology. I will offer some criticisms of the former and defend some aspects of the latter. To this end I will explain why I believe he overstates the metaphilosophical significance of x-phi's results. Nevertheless, I will argue that we can and should keep many of the insights and good methodological habits that come with x-phi and that we can do so without thereby committing ourselves to any metaphilosophical position. X-phi can provide guidance with respect to self-critical philosophical inquiry but it should also be understood within its proper bounds.

1. Intuitions and Institutions

The most important effect of x-phi has been to encourage philosophers to think more carefully about the methodological role of intuition and to maintain a more critical and sophisticated attitude towards the use of thought experiments and examples in philosophical argument. X-phi began in the early 2000s as a movement that deployed empirical methods borrowed from experimental psychology (Knobe 2007, Knobe and Nichols eds. 2008, Stich and Tobia 2016). Its practitioners distinguished themselves from what they disparagingly called *armchair philosophy* by incorporating experimental techniques into debates about the role of intuition in philosophical reasoning (Alexander 2012, Alexander and Weinberg 2007). X-phi was thought to show that many of the premises underlying prominent arguments in the analytic tradition rest on assumptions that are more controversial and parochial than had been assumed.

Thought experiments and more specifically a particular application of thought experiments known as *the method of cases* do not necessarily provide the kind of evidence that philosophers had supposed (Horvath 2015). X-phi has shown that the way people engage with these cases is influenced by demographic factors and framing effects in ways that philosophers had assumed should be irrelevant. Machery and his colleagues have shown experimentally that demographic effects influence how people respond to some of the more famous cases (Machery 2014, Machery et. al 2004, Systma et al. 2015) and he summarizes these results in his book.¹ Similarly, presentation effects can cause us to respond to cases very differently. For example, the order in which cases are presented influences the way that people respond to those cases. It is quite striking for example that in the trolley cases, if one is initially exposed to the footbridge case it influences how one reacts to the switching cases and vice versa. We have long known about the influence of heuristics, priming, and other biasing effects on reasoning tasks. In recent years x-phi showed us that our methods should take our psychological dispositions and limitations into consideration. This has been an important contribution to philosophy.

X-phi should be understood as a reaction to some of the more extreme methodological tendencies in analytic philosophy. In the second half of the Twentieth Century the content and culture of academic philosophy in North America and The United Kingdom was shaped by the work of philosophers at a small handful of prominent departments. With some exceptions, these departments were situated in universities serving the political and economic elite of their societies. The development of the methodology of analytic philosophy during the post-war period is connected to the

¹ Experimental evidence suggests, for example that in the famous trolley cases, millenials are more likely to push the fat man off the bridge than gen x-ers! See Rehman, S. and Dzionek-Kozłowska, J. (2018).

institutional structure of the discipline in ways that merit more careful study than has been undertaken to date.²

In spite of the relative paucity of research into the relationship between the sociology of philosophy and its methodology, there are reasons to posit a connection between the role of notions like *intuition* in philosophy and the institutional structure of the discipline. While we can admit that the sociological factors governing the use of such notions are not well understood, the intellectual history is relatively clear.³ The ways things seemed to philosophers, so-called *intuitions*, came to be understood as grounding philosophical debates and as shaping the acceptable content of those debates (Pust 2013).

This is no longer an automatic assumption. In the 2000s philosophers actively debated the legitimate role of intuition in philosophical inquiry rather than uncritically resorting to intuition talk in the course of philosophical argument. The increasingly critical reflection on intuition talk is due in large part to the influence of x-phi for reasons that I will explain below. It should be noted that not everyone agrees that intuition really played an important methodological role in philosophy. Herman Cappelan, for example, (2012) argued that analytic philosophy relies much less on intuitive evidence than we tended to believe. Cappelan provided reason to believe that the evidential role of intuition is exaggerated in the metaphilosophical literature and that many of the paradigmatic cases of intuition-driven philosophical argument can be reconstructed

² This lack of attention contrasts sharply with the substantial body of research on the development of the philosophy of science in the second half of the Twentieth Century. For example, the social and political role of *paradigm* in philosophy of science has been discussed extensively, see for example Fuller (2000) and Reisch (2019). The lack of sociological and historical attention that a concept like *intuition* has received is striking however, see Symons (2004) for an overview of the changing character of philosophical methodology in the Twentieth Century with special attention to the role of common sense and intuition.

³ The story of how intuition came to the fore in late Twentieth Century philosophy is fascinating. Hintikka (1999) and Symons (2008) trace the blend of ordinary language philosophy, Chomskian linguistic methodology, and the notion of commonsense, in such canonical venues as Kripke's *Naming and Necessity* (1980).

without recourse to intuition. Regardless of whether Cappelan's account of the relatively minimal evidential role of intuition is correct, intuition talk was certainly part of the rhetoric surrounding philosophical practice well into the early 2000s and it remains a prominent feature of debates in moral philosophy.

The critical attitude towards the authority of elite intuition that emerged in the 2000s is connected with the emergence of a critical attitude towards the authority of elite institutions. Snobbishness in Anglo-American philosophy manifested itself in a variety of ways, but having the right sorts of intuition certainly seemed important for membership in the club. For example, it was common for philosophers to dismiss the un-pedigreed as having a *tin ear* for philosophical questions and problems. To 'have a tin ear' meant something like having a deficit with respect to the ways that philosophically significant topics should correctly seem. The notion of the philosophical *tin ear* was only occasionally mentioned in print but many of us recall its being part of the informal disciplinary conversation well into the late 1990s. Judging who possessed a tin ear, who truly 'saw to the heart of philosophical problems' etc. was thought to be a role for gatekeepers, and recommendation letter writers of various kinds.⁴ If taken seriously, these judgments can play a role in determining who ought to count as the appropriate consumers and producers of philosophical research as well as the appropriate content of

⁴ It is rare to find the phrase 'tin ear' being deployed against opponents in published text. Jonathon Weinberg collects some instances and discusses the significance of what he jokingly calls *tinnitus philosophicus* in his (2016 299-301). It is notable that analytic philosophy's gatekeeper-in-chief, Brian Leiter uses the phrase in his criticism of non-philosophical interlopers. Here's an example where he complains about law school faculty trespassing into the philosophers of law: "Academic lawyers who tried to intervene in some ways made matters worse in virtue of having a tin ear for philosophical questions and problems." (Leiter 2006)

that research. Happily, talk of tin ears and shared intuitions has considerably less influence in contemporary philosophy. As Helen Beebe notes, we now recognize that “there is no independent way of establishing which of the rival intuitions is the ‘wrong’ one: one cannot, unfortunately, make an appointment with the intuition-equivalent of an optician and take a tin-ear test.” (2013).

The rise of x-phi coincided with large numbers of philosophers actively and publically discussing the way that social injustices manifested themselves within the institutional structures of the profession. While feminist philosophers and non-white philosophers had made these criticisms in earlier decades, it became common by the late 1990s to remark on unfair aspects of the demographics, and culture of our discipline. Meanwhile, beyond philosophy’s borders, exciting developments were afoot in the biological sciences, particularly in neuroscience and genetics. These developments encouraged some of us to believe that the project of naturalizing large parts of traditional philosophy was feasible and within reach.

By the late 1990s change was in the air. Some of us sought closer contact with the natural sciences, some of us cleaved to the mathematical precision of formal methods, and some of us tried to engage with social and political realities that were not being tackled in ideal theory. Furthermore, by the late 1990s it was increasingly common for philosophers to have had at least some graduate-level education in disciplines outside of philosophy, for example in psychology, the natural sciences, computer science, or

mathematics. For all these reasons, there was a receptive audience among my Gen X peers for the empirical methods and anti-elitist attitude of the x-phi pioneers.⁵

2. The role of the psychology of reasoning in philosophical methodology

The history of philosophy is regularly punctuated by periods where metaphilosophical reflection takes center stage and becomes an explicit concern for philosophical communities. At these moments, philosophers have tried to understand both the scope and limits of philosophical inquiry and its status relative to the natural sciences and other forms of inquiry. Philosophers also concern themselves with the usefulness or harmfulness of philosophy, as well as its political and social function. X-phi differed from the usual forms of self-critical reflection by introducing the empirical study of philosophical practice.

A focus of x-phi research has been the so-called method of cases. Critical engagement with the method of cases serves as the basis for some of Machery's broader metaphilosophical claims. However, as Margot Strohminger (2018) has pointed out, many of the central texts in recent metaphysics and philosophy of language simply do not rely on the method of cases. While this blunts the generality of Machery's metaphilosophical criticisms, the method of cases has nonetheless played an important role in the history of recent analytic philosophy, especially, perhaps in the pedagogical

⁵ Of course, x-phi was only one of a host of developments that pushed the discipline in new directions. Feminist philosophy, formal philosophy, comparative philosophy, philosophy of race, and many other sub-disciplinary innovations flourished as supplements or alternatives to what philosophy had been, well into the late 1990s.

practices of academic philosophy. For example, many of us use the method of cases as a way of introducing problems and questions to our students. That approach is flawed as we shall see. X-phi presents experimental evidence that demographic factors and presentation effects influence our interpretation of philosophically relevant aspects of prominent cases in the literature. X-phi also shows significant divergence between the ways that philosophers and non-philosophers interpret these cases.

Let's begin by explaining what the method of cases is before examining the criticisms presented by experimental philosophers. The method of cases is built around a narrative describing an actual or a possible scenario. These are usually cases where the content is weird or exotic in some way and the narratives are engineered for philosophical purposes. The most widely discussed include, for example, trolley and footbridge cases where we reflect on deontological and consequentialist approaches to the question of whether or how to sacrifice one life for many. In metaphysics there are fission and fusion cases like the Ship of Theseus problem where we are encouraged to reflect on the notion of individuality. In epistemology, there are the standard Gettier cases, where we consider the necessary and sufficient conditions for knowledge. In all such cases, extensive narrative setting involves detail that frames and moves our thinking about the core issues. These narratives work to direct our focus on particular questions and to pull apart aspects of the phenomenon of interest. For example, in the Gettier cases, we are asked to consider the nature of knowledge in a case where justification and truth are pulled apart. These properties are usually bundled together in ordinary life. By using cases, philosophers strive to improve on what Michael Strevens (2019) calls our *starter concepts* of notions like knowledge, personal identity, moral responsibility, and the like.

By contriving cases that distinguish various aspects of the concept under consideration, one way of understanding the purpose of the method of cases is as part of a project of identifying what is absolutely necessary to the relevant concepts. For example, the Gettier cases gave philosophers reason to think that knowledge is not *really* justified true belief since the thought experiments show that the justification and the truth of a proposition can be considered separately.

The method of cases is not always deployed as a way of sharpening our concepts. As we shall see below, in the case of Searle's Chinese Room thought experiment, it is precisely the vagueness of the starter concept of *understanding* that Searle relies upon to support his argument. Nevertheless Strevens is right to point out that the method of cases is intended to put readers in an artificial context which forces focus on specific philosophical questions.

The method of cases can be understood as a strategy in which philosophers engineer ways to block ordinary strategies or habits for thinking through problems as they arise in daily life. For example, when initially introducing the trolley problem to students, instructors inevitably encounter very sensible student reactions where practical solutions to the problem of an impending trolley accident are offered. Students might suggest shouting a warning to the people idling on the dangerous tracks, interfering with the mechanical operation of the trolley or the tracks, etc. In response to untutored practical responses and alternatives, the instructor must engineer the case such that the practical interventions and alternatives are not available and the ethical issue is made unavoidable.

In ordinary life, as Strevens notes, our starter concepts usually suffice. They are good enough to resolve the kinds of practical problems that we are likely to encounter. But as philosophers we realize that our starter concepts of knowledge, personal identity, moral responsibility, and the like are defective in ways analogous to our pre-theoretical concepts of gravity, light, or time. Our starter concepts fail in ways that are similar to the ways that our folk physics fails in extreme or exotic scenarios.

Our starter concepts work within the parameters of normal usage. Philosophers recognize that pragmatic success in a narrow domain is not necessarily a praiseworthy epistemic achievement. It might simply indicate that our starter concepts have lived a very sheltered life where they have not had to encounter any significant challenge. One purpose of the method of cases is to help us to progress beyond such pragmatically- or culturally- certified starter concepts. In the exotic scenarios, as the philosopher presents them, the reader is unable to jump onto the trolley or to push the potential victims out of the way. By blocking the ordinary ways of solving problems, we are encouraged to think about what the philosopher believes is essential to the problem. Or at least this is the way things appeared before x-phi.

It is difficult to disagree with Strevens' view that we should cultivate a critical posture towards our starter concepts. However, some of the results of x-phi show that the method of cases is not always the best way to achieve that end. One problem with traditional attempts to engineer cases is that they involve asking audiences to reason in abnormal contexts. X-phi draws on a long tradition in psychology of showing how unreliable or systematically faulty our reasoning can be in such abnormal contexts. Even apparently simple reasoning tasks can be made difficult when framed in unfamiliar or

abnormal ways. Peter Wason, in the 1960s showed how difficult even simple inferences like modus tollens turn out to be in an unfamiliar or artificial context (1968). By contrast, human reasoning is generally reliable when anchored in familiar contexts (Fiddick et.al 2000). Perhaps most famously, Tversky and Kahneman in the 60s and 70s showed how unreliable our probabilistic judgments are, even for mathematically educated subjects, in certain kinds of artificially constrained experimental contexts (Kahneman et. al 1982).

It would be a mistake to conclude from the work of Tversky and Kahneman that mathematically trained subjects do not really understand probability theory. It would also be a mistake to read the Wason cases as demonstrating that people cannot understand modus tollens. To take the position that people simply cannot reason correctly is both self-undermining, and runs contrary to the fact that we successfully reason about both probability and logical inference in a wide range of cases. Instead, these results are best regarded as indicating that when we reason in unfamiliar contexts, we often reason poorly, we resort to familiar habits of thought and stereotypical judgments that can lead us to make errors.

The method of cases involves carefully engineered narrative contexts that can only work by blocking the kinds of normal anchors that familiar contexts provide. Perhaps it shouldn't surprise us to find that we are not very good at reasoning about exotic cases that are divorced from situations we ordinarily deal with. When things get strange we tend to fall back on familiar or easy patterns of normal reasoning.

Notice that what counts as normal reasoning will vary depending on what one's normal life has served up so far. To some extent this will result in demographic differences. X-phi showed experimentally that people who grow up in Chinese societies

tend to have different starter concepts with respect to knowledge, modal concepts, or concepts concerning moral responsibility than people who grow up in The United States. While X-phi has convincingly demonstrated some of these demographic differences, cultural anthropologists had already offered a wealth of interesting data concerning cross-cultural variation with respect to philosophically significant concepts. X-phi confirmed what anthropologists had long argued, namely that many of our most important starter concepts vary demographically.

In the Twentieth Century, anthropologists showed us that cultural variation with respect to concepts like knowledge can be dramatic. For example, in his ethnographic study of the Ilongot people in the Philippines, Renato Rosaldo notes that the Ilongot only report knowing what they have witnessed themselves (Rosaldo 1980). Why the Ilongot hold this version of the starter concept of knowledge can be understood in relation to their cultural practices related to status, experience, and gender. As Rosaldo explained, in Ilongot society the ability to recount personal experiences of adventures and travel to distant places were held in high regard. Such personal experiences were the domain of Ilongot men rather than women. Claiming *to know* and claiming to be *acquainted with* were thus closely entangled given the manner in which the Ilongot apportion social status.⁶

Anthropology has had relatively little influence among mainstream analytic philosophers. If it had, the fact that ordinary or starter concepts of knowledge and other philosophically interesting notions vary among different demographics would not have been controversial or surprising. One uncontroversial but important lesson of x-phi is that people come to philosophy with very different starting points and that these starting

⁶ Thanks to Kareem Khalifa for stimulating discussion of Rosato's ethnography

points influence how we think about controversial cases.⁷ It is important that our philosophical methods and especially our pedagogical practices be sensitive to these differences.

As we saw above, Strevens argued that the method of cases serves as a way to critically interrogate our starter concepts. However, psychological evidence shows us that we find it difficult to reason well about extreme or unusual cases. Because of this, the method of cases might generally fail to teach and persuade. The contrived and artificial character of many of the cases is likely to encourage subjects (at least at first) to resort to crude and oversimplified patterns of thought in their own reasoning. This explains why philosophers diverge dramatically from non-philosophers in their judgments with respect to cases as Machery discusses. For philosophers, these examples are not exotic, but are instead shop-worn tools of the trade.

3. Suspending judgment about cases vs. critically engaging with their assumptions

Even for experts, the method of cases can fail to lead to consensus. In fact, Machery's skepticism concerning the usefulness of the method of cases is focused precisely on examples where expert disagreement – disagreement among philosophers – seems intractable. His conclusion is that in those cases, philosophers should suspend judgment on the specific question under consideration (2016 127). For example, given the apparently intractable disagreement about the possibility of philosophical zombies

⁷ Notice that we could say similar things about all branches of inquiry.

philosophers should simply abandon that debate and move on to other problems.

However, in this section I will argue that this is not the only metaphilosophical path open to those of us who take the experimental insights of x-phi seriously. Rather than recommending that we simply give up, we might take the lessons of x-phi as resources for thinking more carefully about the manner in which cases are engineered.

The source of intractability need not be the method of cases per se. It might, for instance result from the manner in which a standard case or thought experiment has framed a specific philosophical problem. X-phi encourages a critical spirit with respect to the way that cases frame philosophical questions. In this spirit one could imagine that alternative cases or even an as yet undiscovered refinement of an existing case might render the problem tractable. Indeed, one could imagine experimental philosophers taking an interest in empirically investigating such strategies.

Consider taking this critical posture towards, for example, Searle's famous Chinese Room thought experiment (1980 417-418). Searle's argument is one of the most influential uses of a thought experiment in late 20th Century philosophy. Machery regards it as a "striking example of the method of cases" (2016, 23). It purports to show that machines implementing programs cannot understand natural languages. There is a range of conflicting reactions to Searle's argument and it has spawned a large literature of responses. It can serve as an example of the use of the method of cases that generates intractable disagreement and it likely counts as an instance of the kind of problem where Machery would urge us to suspend judgment. However when we are asked to determine whether the Chinese Room understands Chinese, suspending judgment is premature and perhaps not properly in the spirit of x-phi for reasons I will explain below.

Searle's famous case (Searle 1980), which Machery reprints in Chapter One, leads readers through roughly the following chain of reasoning:

1. I, the reader, understand what is meant by 'understand Chinese'
2. The imagined Chinese Room contains the resources (a book of instructions written in English for a non-Chinese speaker) to permit the person in the room to respond in writing in meaningful ways to questions posed by Chinese-speaking interlocutors. Those interlocutors will judge the responses provided by the Chinese Room to be meaningful sentences of Chinese.
3. The person following instructions in the Chinese Room does not understand Chinese
4. The computers that are in use in artificial intelligence are like the Chinese Room, they are nothing but systems of rules whose syntax can serve as the basis for mechanical implementation of those rules.
5. Therefore, computers do not and cannot understand natural language.

The Chinese Room argument is a piece of philosophical sleight-of-hand. The narrative supports Searle's argument only if the reader fails to notice that Searle is being too generous with some aspects of the case and too stingy with others at the same time. Searle flatters the reader by granting full credit to their starter concept 'understand'. This is too generous. The folk or starter concept of understanding is simply inadequate to the task of determining whether computers are capable of understanding natural language. In this sense, we can sympathize with Machery's impulse to simply walk away

from the inevitable mess that the underdeveloped folk notion of understanding generates in Searle's case. However, there is another path open to us. We can take a critical attitude towards the terms in which the thought experiment is framed. Such a critical perspective would recognize that Searle is leveraging our ordinary sense of when it is appropriate to say that someone understands a language. As I will explain below, this loose sense of 'understanding a language' is not good enough to answer the question of whether machines can understand natural language. One way to respond to the case is to recognize that the starter concept of understanding used here needs to be refined.

Searle asks us to imagine that the human being in the Chinese Room is not a Chinese speaker. Thus, the human being in the box straightforwardly fails to understand the language in the ordinary, starter-concept sense of *understand*. Since there is no other person in the Chinese room, we are inclined to judge that there is no understanding of Chinese taking place in the room. At the same time, we recognize that the room as a whole passes any ordinary behavioral tests that we might set for it as we try to determine whether we are communicating with a Chinese interlocutor. This is the source of the intractable disagreement.

Notice that the starter concept of understanding that is at work here ordinarily applies to human persons. The conditions for applying the concept are vague. Given our starter concept of understanding, one can be said to understand English even when one does not understand some large subset of English sentences. So, for example, it is appropriate to say that a person understands English even though they might not understand many sentences in Shakespeare's plays or many sentences of quantum field theory. And yet, if someone is limited to 'restaurant French' we might say, with some

justification, that they do not *really* understand French. The threshold for making this judgment is not well-defined. Perhaps this is because our starter concept of ‘understanding a language’ is used in contexts that do not require a precise number of sentences below which we would not count someone as understanding the language. What starter-concept-users mean when we talk about ‘understanding Chinese’ is good enough for all the ordinary occasions in which we would normally need to think about such things.

The Chinese Room is not an ordinary context. Perhaps if we first sharpened the starter concept of ‘understanding’ we would be in a better position to address the problem but Searle’s argument presupposes that our ordinary concept of understanding can do the work required of it in this context and more importantly that our starter concept correctly serves as our standard for judging the case.⁸ In simple terms, we must grant step (1) in the chain of reasoning in order for Searle’s argument to proceed.

X-phi can be understood as providing tools for us to begin critically assessing, or at least being sensitive to variation with respect to our starter concepts. Someone who learned the lessons of x-phi regarding demographic variation and the slipperiness of intuitive starter concepts would insist on precision with respect to the notion of ‘understanding’ being used in the thought experiment. When asked whether the Chinese Room understands Chinese, the best response is not to suspend judgment but to say: “It depends what you mean by ‘understand’.”

⁸ As we have seen above, philosophical inquiry generally requires a critical attitude towards our starter concepts. Ordinary Language Philosophy, a tradition of philosophers who were at the peak of their powers in the 1950s and early 1960s, was a historically influential exception. Philosophers like Ryle, Austin, the later Wittgenstein and others did not hold this critical attitude towards starter concepts. Instead, ordinary uses of terms like ‘real’, ‘belief’, and ‘understand’, were seen as standards against which philosophical uses of those terms should be judged. It is notable in this context that Searle studied with John Austin and that his own thought was influenced by some of the methodological principles of ordinary language philosophy.

Searle's generosity with his readers' understanding of 'understanding' contrasts with his stinginess with respect to the details of the thought experiment itself. Searle gives almost no account of the algorithm for instructing the non-Chinese speaker in how to respond to Chinese questions. He asks us to imagine that a book has been produced with rules governing all the relevant kinds of questions and responses that the Chinese room will face. Here, critical readers should ask for more details before suspending judgment. As usual in the method of cases, Searle must declare a limit to his audience's demand for details. As we saw in the discussion of Strevens' view of the method of cases, the philosopher must limit the detail of the narrative in order to close off certain lines of questioning by fiat. In this case, in order for his argument to be persuasive, readers must trust Searle that his hypothetical computer scientists figured an algorithm for competent use of natural language and that this algorithm could be converted to the form of a book of instructions.

As we have seen, Strevens' defense of the method of cases discussed above does not apply to the Chinese Room case. This is because Searle's thought experiment is not helping us to sharpen our starter concept of *understanding*. On the contrary, Searle is leveraging the weaknesses of the starter concept and using it to help persuade his readers that artificial intelligence is doomed to fail. Searle's Chinese Room is a thought experiment that is devoted to supporting a particular philosophical conclusion while leveraging a defective starter concept of understanding. Far from helping us to sharpen or improve that concept, Searle's thought experiment relies on its deficiencies.

I've taken the time to discuss the Chinese Room in order to indicate what I take to be an alternative way of responding to x-phi's treatment of the method of cases. X-phi

arms us with the following insights: We have considerable evidence to suppose that many of the standard cases that philosophers have used are of relatively limited pedagogical value for reasons discussed above. It is quite possible that the method of cases can serve as a useful means of discovering truths about our concepts, but as x-phi has shown, it is not a reliable method for justifying those discoveries to broader audiences. Given the lessons of x-phi, readers will be equipped to approach the Chinese Room critically but they need not simply suspend judgment. Instead, readers will demand conceptual clarity and appropriate levels of detail.

4. Modal immodesty and Metaphilosophy

Machery's metaphilosophy is naturalistic and pragmatic in spirit. In this section I will argue that he is mistaken to regard x-phi's critical engagement with the method of cases as providing significant support to this broader metaphilosophical position.

Let's begin with his criticism of modal reasoning. Our modal reasoning is concerned with what might, must, or could not be the case. To varying degrees, we all assume that a world other than the world we know is possible. As philosophers we ask, for example whether it is possible for the mind to exist without the body? As mathematicians we ask whether it is *necessarily* true that the square root of nine is three? In ordinary life we reflect on what we are morally obliged to do in various situations and we try to understand what kinds of decisions or lives are open to us. We use imagination and construct fictional narratives. For example, we consider what would it be like to deny certain features of the actual world and consider what things would be it like if, for instance, I had actually become an electrician instead of a teacher.

While much of this kind of reasoning would be acceptable to Machery, he thinks that philosophers have gone too far. Machery opens his book by arguing that philosophers' "flights of fancy" (2017,1) are curbed by the work of x-phi. The criticism of the method of cases is taken to support his skepticism towards modal reasoning. However, it is unclear how the two are connected. As Margot Strohmingher has noted, for example, there is a range of distinct paths to modal knowledge that have been discussed in the recent literature (2018). Most importantly, the so-called inferential path to modal knowledge exemplified by Kripke's arguments in *Naming and Necessity* (1980, 159). As Strohmingher (2018) explains, the inferential path takes the following form:

Given that one knows that p and p is necessary if true, one can infer that p is necessary.

The inferential path to modal knowledge is unrelated to the method of cases and therefore evades Machery's criticism.

Of course, Machery recognizes that the experimental results he presents are not strong enough to block the *possibility* of modal knowledge. However, the purpose of his argument is to simply to persuade us that we should be less confident than we suppose that we have the kinds of access to modal knowledge. In practice, epistemic humility is a virtue and I have described the salutary effect of x-phi in this regard. However, there is a difference between adopting self-critical methodological practice and endorsing metaphilosophical prohibitions.

Take a metaphysical claim where we all seem to agree. We say, for example that it is *impossible* for something to be red and green all over at the same time. We say also, for example that it is impossible for me to imagine that I own a seven-sided regular polyhedron. Here, we do seem to have some knowledge of what must be the case. Sometimes philosophers worry that this knowledge does not compare favorably with knowledge we have of the actual world. And yet, in some sense claims like these have a significantly higher epistemic status than the particular claims of the empirical sciences. My experiments can go wrong, but at least some of my metaphysics will be unshakeable.

Machery is unmoved by arguments of this kind and argues instead, on largely pragmatic grounds, for the sterility of much of contemporary philosophical reasoning. In its place he proposes an alternative account of philosophical practice that I will describe below.

Before turning to his proposed alternative it is worth considering how difficult it is for those with pragmatist leanings to engage in metaphilosophical argument. Let's assume that such arguments can only take place in the context of the existence of competing views of philosophy. Let's also assume the goal of a metaphilosophical argument concerning philosophical method, for example, is to show that one's opponent is either operating with a bad method or that their method is not a philosophical method. In the first case, we would be involved in showing that one method is better than another. Here, judgments of better or worse rely on reference to some standard. In Machery's case, this standard is pragmatic. Adopting both a pragmatic metaphilosophy and a pragmatic standard for judging the relative value of different metaphilosophies risks circularity. This will not be a problem for the pragmatist until they attempts to argue for

their metaphilosophical position. They might claim, for instance that the proof of the value of the pragmatist pudding is its successes. However, this will not work as a convincing defense of a metaphilosophical position like Machery's that advocates suspension of judgment along the lines discussed above. This is because it is difficult for the pragmatist to argue that suspending judgment counts as success with respect to some philosophical problem. It is hard not to see the pragmatist as encouraging us to simply abandon philosophical debates rather than successfully resolving them.

Machery advocates an alternative approach to philosophical practice loosely modeled on Carnapian explication which he calls *naturalized conceptual analysis*.⁹ The goal of Machery's enterprise is to identify flaws with existing concepts and to engineer or reengineer concepts for specific purposes. Where his approach differs from traditional conceptual analysis is in its emphasis on what he calls "empirical invalidity". Concepts are empirically invalid on this view when they do not reliably serve in inquiry. In Chapter 7 he provides an analysis of the concept of innateness as an example of what he has in mind.

He explicitly identifies this project with engineering and follows Gramsci in seeking to remedy morally or politically flawed concepts. According to Gramsci, the bourgeoisie maintains illegitimate power over the rest of us in virtue of a cultural hegemony where ideology, rather than physical force organizes social hierarchies by making the established order seem normal or commonsensical. Gramsci's account gives intellectuals a central role in the construction of social reality insofar as they actively contribute to and sustain the ideological environments that shape our lives (Gramsci 1982). This approach can be challenged on a variety of levels. Most obviously however,

⁹ For Carnap's account of explication see the first chapter of his *Logical Foundations of Probability* (1950)

it assumes some prior settled moral or political orientation that guides the engineer's ameliorative project. Machery refers in passing to critical theory as one of the remaining tasks for philosophy. However, if we were to settle on an orientation such as Gramsci's or that of later critical theorists we would presumably have arrived at our position via a method other than conceptual engineering. It is difficult to imagine arriving at a plausible ethical or political position without engaging in precisely the kinds of philosophical inquiry that Machery criticizes elsewhere. Historically, for example, the tradition of critical theory assumed a broadly Marxist political orientation. One certainly would not want to uncritically assume the truth of a Marxist metaphysical and normative framework when embarking on one's putatively ameliorative project.

Let's try to distinguish the kind of definitional project involved in Machery's conceptual engineering with a more traditional view of philosophy. Consider the following: Let's say P wants to invite some of her colleagues from work to her home for a social event. She doesn't want to call this event a 'party' because that term has connotations that are unsuited to what she has in mind. She wants to be hospitable, but she would prefer not to see her colleagues drunkenly carousing, dancing, or talking too loudly. She wants to improve her working relationship with her colleagues, but she is less interested in making new friends and more interested in the smooth and efficient operation of her place of work and her own success therein. In this context, P is a pragmatist whose purposes are clear. As she drafts the invitation email to her colleagues, she settles on calling the event she plans a 'get-together'. In her invitation she describes (defines) what she means by 'get-together' in a way that she hopes will convey the scope and limits of the event that she intends to host. P's definition of 'get-together' as she

crafts her email message involves the kind of conceptual engineering project that Machery has in mind. It has an empirical element insofar as the project might fail. Perhaps she is insufficiently clear about what the characteristics of a get-together are, or perhaps her mention of wine in the invitation gives some of her guests the wrong impression. If so, she will know better next time and will revise her description/definition accordingly.

It would be strange to think that what P is doing should be called ‘philosophy’. To begin with she is not interested in discovering what constitutes a party or a get-together. Her goal is not knowledge or increased understanding, but is, something else, perhaps professional advancement or social harmony in the office. From P’s perspective, it is certainly the case that the question of knowing what really counts as a party or get-together is at best orthogonal to her purposes and probably a pointless waste of her time. There are more pressing matters to attend to. When philosophers take this attitude towards the philosophical enterprise we generally call them pragmatists.

People who want to know what parties really are have philosophical rather than pragmatic motivations. Not everyone who is a philosopher wants to know what a party really is, but if you are concerned with knowing such a thing, you are probably a philosopher and probably not a pragmatist. Now, it is likely that there is nothing more to being a party than the social conventions involved. Perhaps there is no essence of party. But it is important to notice that discovering whether that is true would involve a philosophical rather than a definitional practice.

Philosophical practice takes many forms. Nevertheless, almost all philosophers value texts or spoken presentations that persuade audiences to move from obviously true

premises to surprising conclusions. We disagree as to how we ought to arrive at these premises, the kinds of persuasion that are appropriately philosophical, and the subject matter or content of genuine philosophy. Nevertheless, we all agree that something like a movement from the obvious to the surprising is a feature of philosophy. We can call this the inferential view of philosophy (setting aside the need to be clear about what an inference is for now) and contrast it with a view that sees philosophy as a conceptual engineering project.

In the engineering project there are surprises, but the source of these surprises is limited. Our engineered projects can fail in surprising ways in virtue of the way that the world turns out to be. These surprises can fall into two broad categories of correction. We might be surprised to learn that nature has features that foil our plans. Like the builders of the doomed Tacoma Narrows suspension bridge, properties of the natural world that we had failed to take into consideration can present themselves. But we might also learn through practical failure that the norms and presuppositions governing our plans are incoherent. We might, for instance, have misguided or evil goals. Utopian projects of social engineering have a history of ending in failure. This failure is not always because the natural or social world fails to bend to the will of the social engineer. Sufficiently powerful agents can certainly create new social orders for a time. The failure of utopian projects can arise because the motivations of the social engineer are generally self-undermining, immoral, or incoherent. Critical interrogation of the kinds of norms, concepts, and presuppositions governing engineering are not necessarily part of engineering itself.

Conclusion:

I have emphasized that there is a difference between x-phi itself and the metaphilosophical lessons that one might draw from x-phi. Notice that one can be an appreciative consumer of the results of x-phi without buying naturalistic or pragmatic metaphilosophy. One might share its critical posture towards philosophical methodology while differing with the broader metaphilosophical claims in Machery's book.¹⁰ If one is not already convinced of pragmatism or naturalism, the discoveries of x-phi are unlikely to make too much difference to one's metaphilosophical position.

The accomplishments of experimental philosophers have been important and we should be appreciative of the changes that Machery and his colleagues have helped to bring about in our discipline. Philosophy is a more self-critical, open, and interesting enterprise than it was 20 years ago and this is partly due to the salutary influence of x-phi.

Acknowledgments:

This work was partly supported by NSA Science of Security initiative contract #H98230-18-D-0009.

¹⁰ For example, I have criticized the use of intuition in analytic philosophy (2008), and have advocated a more formal approach to the methodology of philosophy than is commonly held by our colleagues (2005; 2006), and I have argued for a naturalistic approach to some central questions in philosophy of mind (2012). My impression of changes in our discipline lead me (unscientifically) to the opinion that x-phi has improved the methodological hygiene of our discipline, but perhaps only a proper empirical study could tell conclusively.

Bibliography

- Alexander, J. 2012. *Experimental philosophy: An introduction*. Polity.
- Alexander, J., & Weinberg, J. M. 2007. Analytic epistemology and experimental philosophy. *Philosophy compass*, 2(1), 56-80.
- Beebe, H., 2013. Women and deviance in philosophy. in K. Hutchison & F. Jenkins (eds.) *Women in philosophy: What needs to change*, pp.61-80. Oxford: Oxford University Press.
- Cappelen, H., 2012. *Philosophy without intuitions*. Oxford: Oxford University Press.
- Carnap, R., 1950. *Logical foundations of probability*. Chicago: University of Chicago Press
- D'Oro, G. and Overgaard, S. eds., 2017. *The Cambridge companion to philosophical methodology*. Cambridge University Press.
- Fiddick, L., Cosmides, L. and Tooby, J., 2000. No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition*, 77(1), pp.1-79.
- Fuller, S., 2000. (2000). *Thomas Kuhn: A Philosophical History for Our Times*. Chicago: University of Chicago Press.
- Gendler, T.S., 2010. *Intuition, imagination, and philosophical methodology*. OUP Oxford.
- Gramsci, A., 1982. *Selections from the prison books*. Trans by Q. Hoare and G. Nowel Smith. London: Lawrence and Wishart.
- Haug, M.C. ed., 2013. *Philosophical methodology: the armchair or the laboratory?* Routledge.
- Hintikka, J., 1999. The emperor's new intuitions. *The Journal of Philosophy*, 96(3), pp.127-147.
- Horvath, J., 2015. Thought experiments and experimental philosophy. In Daly, C. ed. *The Palgrave Handbook of Philosophical Methods* (pp. 386-418). Palgrave Macmillan, London.
- Hendricks Vincent, John Symons. 2005. *Formal philosophy*. Automatic Press.

- Hendricks, Vincent, John Symons. 2006. No mere formality. *The Philosophers' Magazine*, (35), pp.35-38.
- Hume, David 1907 [1742]. *Essays: Moral, political, and literary* (Vol. 1). Longmans, Green, and Company.
- Kahneman, D., Slovic, S.P., Slovic, P. and Tversky, A. eds., 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.
- Knobe, J. 2007. Experimental philosophy. *Philosophy Compass*, 2(1), 81-92.
- Knobe, J. and Nichols, S. eds., 2008. *Experimental philosophy*. Oxford University Press.
- Kripke, S., 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Leiter, B., 2006. Introduction: From Legal Realism to Naturalized Jurisprudence. *University of Texas Law, Public Law Research Paper*, (104). Available at SSRN: <https://ssrn.com/abstract=926079> or <http://dx.doi.org/10.2139/ssrn.926079> (last accessed October 4, 2019)
- Machery, E., Mallon, R., Nichols, S., & Stich, S. P. 2004. Semantics, cross-cultural style. *Cognition*, 92(3), B1-B12.
- Machery, E. 2014. What is the significance of the demographic variation in semantic intuitions? *Current controversies in experimental philosophy*, 3-16.
- Machery, E., 2017. *Philosophy within its proper bounds*. Oxford University Press.
- Pust, J., 2013. *Intuitions as Evidence*. New York: Routledge.
- Rehman, S. and Dzionek-Kozłowska, J., 2018. The Trolley Problem Revisited. An Exploratory Study. *Annales. Etyka w życiu gospodarczym*, 21, pp.23-32.
- Reisch, G.A., 2019. *The Politics of Paradigms: Thomas S. Kuhn, James B. Conant, and the Cold War "Struggle for Men's Minds"*. SUNY Press.
- Rosaldo, R., 1980. *Ilongot headhunting, 1883-1974: A study in society and history*. Stanford University Press.
- Searle, J.R., 1980. Minds, brains, and programs. *Behavioral and brain sciences*, 3(3), pp.417-424.
- Stich, S. and Tobia, K., 2016. Experimental philosophy and the philosophical tradition. In Sytsma, Justin, and Wesley Buckwalter, eds. *A Companion to Experimental Philosophy*. John Wiley & Sons

Strevens, M., 2019. *Thinking off your feet: How empirical psychology vindicates armchair philosophy*. Harvard University Press.

Strohming, M., 2018. *Review of Edouard Machery, Philosophy Within Its Proper Bounds*. Notre Dame Philosophical Reviews. <https://ndpr.nd.edu/news/philosophy-within-its-proper-bounds/> (last accessed Oct 4, 2019)

Symons, J., 2008. Intuition and philosophical methodology. *Axiomathes*, 18(1), pp.67-89.

Symons, J., 2010. Ontology and Methodology in Analytic Philosophy. In *Theory and applications of ontology: Philosophical perspectives* (pp. 349-394). Springer, Dordrecht.

Sytsma, J., Livengood, J., Sato, R. and Oguchi, M., 2015. Reference in the land of the rising sun: A cross-cultural study on the reference of proper names. *Review of Philosophy and Psychology*, 6(2), pp.213-230.

Wason, P.C., 1968. Reasoning about a rule. *Quarterly journal of experimental psychology*, 20(3), pp.273-281.

Weinberg, J. M., 2016. Intuitions. In Cappelen, Herman, Tamar Gendler, and John P. Hawthorne, eds. *The Oxford handbook of philosophical methodology*. Oxford: Oxford University Press.