

Experimental Philosophy Without Intuitions: An Illustration of Why it Fails¹

Herman Cappelen

The experimental philosophy movement started out as a criticism of what it took to be the core of philosophical methodology: the practice of constructing cases (or thought experiments) and then eliciting intuition about those cases. According to this view of philosophical methodology, intuitions are used as evidence for or against philosophical theories. A good theory of knowledge, justice, freedom, or reference (for example) should be consistent with intuitions we have about cases (thought experiments). The experimental philosopher's core idea was this: if philosophy is based on intuitions about cases, we need to check what intuitions people have, what influences those intuitions, and whether they track truth reliably. So they conducted experiments to check on that – hence the name. The results were discouraging. They appeared to show that intuitions about cases varies across subjects and are easily influenced by irrelevant factors. They concluded that the standard methodology was flawed and that this undermines an enormous amount of work done in philosophy.

¹ Thanks to Derek Ball, Josh Dever, Edouard Machery for useful discussions, and to Matt Mckeever for detailed comments and suggestions.

In the first chapter of Edouard Machery's new book he rejects this way of understanding X-phi. He endorses the views of some of the most ardent critics of X-phi. Machery agrees with Cappelen (2012) and Deutsch (2015) that the X-phi movement made a fundamental mistake when it described standard philosophical methodology as relying on intuitions about cases. And he agrees with the critics of X-phi that intuitions play no role in philosophical practice. However, he doesn't agree with Cappelen and Deutsch that this makes X-phi pointless. His book is an effort to show how X-phi without intuitions can still be philosophically significant. It's a bit like how you might think about acupuncture: you might think there's something to the practice of acupuncture even though the classical theory of why acupuncture works is false. If so, you reject the view that the Qi (or life force) is moved around so-called meridians by needle pricks, but nonetheless think that the practice of sticking needles into people can be useful. You reject the theory of acupuncture, but preserve the practice. Machery's attitude towards experimental philosophy is similar: the theoretical scaffolding of X-phi (as we find it in Machery, Mallon, Nichols, and Stich 2004, Knobe and Nichols 2008, Nagel 2012, among many others), is false. The appeal such authors make to intuitions (both when describing their experiments and when criticizing or supporting philosophical practice) is like the appeal to Qi in acupuncture. However, says, Machery, lovers of X-phi should not despair: X-phi without intuitions can have massive philosophical significance.

The book is bold, provocative, engaging, ambitious, and well written. It draws on and synthesizes an extraordinary amount of recent work. It is an important contribution to contemporary metaphilosophy that most readers will learn a lot from. I should immediately add that I say all this despite disagreeing with most of the central claims in the book (that this combination is possible is a nice, disturbing, and somewhat puzzling feature of philosophy).

Machery on X-Phi without intuitions

According to Machery, philosophical practice relies essentially on what he calls 'the method of cases' in which philosophers use judgements about thought experiments as evidence.

Experimental philosophers have said these were intuitive judgements, and then gone on to claim that those intuitions were unreliable. Machery, on the other hand, denies that there's anything distinctive about the judgements philosophers make about thought experiments – they are just ordinary judgements. He calls this view 'minimalism':

...no particular property distinguishes the judgments elicited by philosophical cases from everyday judgments. Judgments elicited by philosophical cases do not possess a phenomenology that everyday judgments do not have: In particular, we do not feel we have to make these judgments. They do not have a distinctive epistemic status: For instance, they are not justified a priori. They do not have a distinctive semantic status: They are not analytic. They do not have a distinctive etiology: They do not express our conceptual competence with the relevant concepts. ... minimalist characterizations of the method of cases do not require the judgments elicited by philosophical cases to share any particular content ... For instance, minimalist characterizations do not characterize them as fast judgments that have not been consciously inferred. Some judgments elicited by philosophical cases could be consciously inferred, and others could be slow. (Machery 2017: 20)

This is where Machery agrees with the critics of X-phi, such as Cappelen and Deutch. However, this leaves Machery with a dilemma:

Dilemma: on the one hand, Machery is not a skeptic: he doesn't hold the view that ordinary human judgments are excessively unreliable. On the other hand, he does want to argue that the X-phi has shown that philosophers' judgements about cases are fundamentally unreliable. How can those two positions be reconciled? If philosophical judgments about cases are just ordinary judgements, and there's nothing generally wrong with 'ordinary judgements', what's so deeply flawed about philosophical judgements?

The answer is that philosophical cases have ‘disturbing characteristics’ and these are source of the unreliability revealed by the experiments done by Machery and other experimental philosophers. The disturbing characteristics are, moreover, essential to philosophical cases: it’s impossible to do traditional philosophy without appealing to cases that have the disturbing characteristics.

Machery’s primary focus is on three disturbing features: (i) the **unusualness** of the cases, (ii) that the cases **pull apart** properties that usually go together, and (iii) that it’s **hard to separate** the superficial content of the cases from their target content.

I turn now to some critical remarks about this part of Machery’s project. Most of my discussion will concern one of the disturbing characteristics: unusualness. I think what I say about this will generalize and if so, it has wide reaching implications for the overall project in the book.

Why unusualness is not disturbing

It is easy to describe unusual (infrequent) cases that are easy to reliably judge. That refutes the claim that infrequency/unusualness is a source of unreliability. Here’s an example like that: there are no orange elephants so we don’t encounter them (and there’s little literature that about them). Nonetheless, we have no problem making various judgements about orange elephants. They can be small, run fast, be sick, scary, etc. It’s unproblematic to attribute, the property of, say, *running* to an orange elephant in a thought experiment.

Machery discusses this very objection (in response to Cappelen 2012: 226). He says the objection fails because situations involving orange elephants isn’t what he has in mind when he talks about unusualness. He says, ‘A lie is a lie, whether it is made in an actual situation or in the world of *The Game of Thrones*’ (2017: 121). Ok, so what is unusualness? In reply we get a sufficient condition: “it is unusual if it describes a kind of situation that we rarely encounter in our experience or from description” (2017: 121). It hard to see how this helps. Here is *a kind of*

situation: the kind that involves orange elephants running. That kind is infrequently encountered in experience or descriptions, but we can reliably describe orange-elephant involving situations. We are just as reliable in our descriptions of these as we are in our descriptions of real grey elephants.

To Machery's credit, he tries hard to engage with this kind of issue (in particular in 3.6.1, and the rest of chapter 3), and there's a second line of reply. In his discussion of funny-colored elephants he says that their color is not 'germane to getting it right' (2017: 121). He says that "whether one is adding chairs or pink elephants is not [germane] when one is adding" (2017: 121). This then makes it super important that we're given an account of germaneness, but Machery has nothing to say about this. On the face of it, the appeal to what is 'germane' is unhelpful. Consider the following thought experiment:

- There are two pink, one orange, and one grey elephant in the room. Then seven orange elephants walk in (and none leave).

Question: "How many of the elephants in the room are orange?" The answer to this is easy and we're reliable in making the judgement, but surely the color of the elephant is germane.

Otherwise I just don't know what 'germane' means. The question, after all, is about orange elephants, so how can color not be germane?

Moving on, I want to outline one additional concern about Machery's claim that unusualness is a source of unreliability. Consider our ability to attribute properties like 'unusual situation' to situations. Machery must assume that we are generally reliable when we attribute that property to situations because he assumes throughout his book that he can reliably make judgements about which situations are unusual (he's confident that the cases philosophers appeal to are unusual – that's a central thesis of his book). So, despite unusualness being a source of unreliability, we can reliably judge that situation are unusual. That's paradoxical if unusualness is a source of unreliability. Maybe Machery thinks that we can reliably attribute 'unusualness' to usual unusual

situations, but that we're unreliable when it comes to unusual unusual situations. This move assumes, without argument, that philosophical cases fall in the former category and leaves unexplained why we would be reliable in judging usual unusual cases (unusualness, after all, is what we're supposed to be unreliable at judging).

Against the Essentiality of Unusualness

I've raised some concerns about Machery's claim that unusualness is a source of unreliability. I turn now to Machery's claim that unusualness is essential to philosophical thought experiments. (Of course, if my first argument goes through, then essentiality wouldn't be a serious concern, but the claim is still worth exploring.) What's the evidence for the essentiality claim? Note that the essentiality claim is extraordinarily broad claim about literally hundreds of thousands of papers, arguments, and claims made by philosophers. We should expect detailed empirical engagement with philosophical practice to substantiate the claim. There is nothing of that. Instead we get unsubstantiated generalizations like the following: "Competing philosophical views ... often give the same verdict about ordinary situations. In order to distinguish them, philosophers *must* consider unusual situations." (2017: 115) As a methodological point, I suggest that detailed engagement with the rich practice of philosophy should be expected of anyone making such broad claims. Machery should engage in detailed investigation of a broad range of philosophical texts and for each of them show that unusualness is essential. To see how hard that would be, recall that Machery must be operating with a technical (and complicated) notion of 'unusualness'. He does not, for example, think that science fiction describes unusual cases. Talking foxes, ghosts, flying dragons, magical spells, invisibility cloaks, and such things do not fall under 'unusual' for Machery (and if they did, it would be a *reductio* of his view). Recall also from the earlier discussion that he'll have to distinguish between usual unusual situations and unusual unusual situations. Here is what I think will happen: Machery can try to come up with a diagnostic for unusualness (or unusual unusualness) in the form of some conditions, C. Then his claim is that philosophers not only do, but must rely on cases with C features. His job then is to show that a few hundred arguments in the philosophies of mind, language, science, art,

mathematics, etc. have to have the C features. The book doesn't even begin to provide such evidence and I'm not an optimist that this would turn out as Machery needs it to.

Before I leave essentiality claim behind, one final point. Note that essentiality is a strong modal claim about the properties of the thing that Machery calls 'philosophy'. It is, in effect, exactly the kind of immodest modal claim that Machery in chapter 6 argues is not within the proper domain of philosophy. Machery is, after all, also doing philosophy in this book and the lessons of chapter 6 should apply to this book as well.

I have focused on Machery's claims about unusualness for two reasons: First, what I have just said about unusualness generalizes to the other alleged disturbing features, e.g., the feature of pulling apart properties that normally come together. Limitations of space forces me to leave that as an exercise for the reader. Second, if you spot me that the points about unusualness generalize to the other disturbing features (a big ask), then it's a serious problem for Machery's entire project. More or less everything that follows depends on these disturbing characteristics really being deep and essential sources of unreliability. If they are not sources of unreliability or not essential, then: (i) we don't have an alternative account of why X-phi is philosophically significant: the disturbing characteristics are supposed to play the role of intuitions in traditional X-phi. They're the new boogeyman. If they're not as bad as advertised (or not bad at all, as I think), it's no longer clear why we should think X-phi without intuitions is philosophically significant. (ii) The aim of Chapter 6 is to delineate what philosophers can and cannot do. The argument goes like this: *Philosophers tend to ask questions about metaphysical necessities. In order to get those answers, we have to rely on thought experiments that have the disturbing characteristics. Our judgements about such cases are unreliable. Corollary: we should stop asking for metaphysical necessities.* Machery says:

In the light of the limitations of the judgements elicited by philosophical cases revealed by the previous chapters, I argue in this penultimate chapter that, even if there are modal facts ... many of the modal facts that matter for philosophy are beyond our epistemic

reach. We philosophers are not in a position to acquire the desired modal knowledge.(2017: 186)

It is clear, therefore, that the core claims of chapter 6 build on the previous chapters: they presuppose that Machery has correctly identified essential sources of unreliability in judgements elicited by typical philosophical cases. If I'm right, that part of the project fails and so the foundations for the claims in chapter 6 evaporates.

Before moving on to some comments on the last chapter of the book, I pause to mention a view that springs quite naturally from these lines of criticism: Metaphilosophers should give up the idea that there's a small set of features that characterize all philosophical cases. I would go further: the term 'method of cases' doesn't denote a theoretically useful class. It encourages the thought that there's uniformity where there isn't. The thing we call 'cases' come in too many varieties. The source of their degree of difficulty is multifaceted and cannot be separated from their specific subject matter. Issues having to do with mathematics, art, action, the mind, language, and morality are difficult for their own peculiar reasons, often specific to particular theories and arguments. Philosophical methodologists should give up the simplifying assumption that there's a unified class of things called 'cases' and that by focusing on a handful of feature of this fictional class, philosophers get to criticize work done by thousands of colleagues without careful engagement with their arguments, theories and texts (for more on how the class of 'cases' tend to mislead metaphilosophers, see Cappelen and Deutch 2018.)

Conceptual Analysis 2.0

I turn now to the final chapter of the book. In the penultimate chapter Machery argues that the kinds of questions philosophers tend to ask cannot be answered. In the seventh and last chapter he gives a partial answer to the question: "So, what should philosophers do?" The proposal is that we should engage in what I will call *Conceptual Analysis 2.0* (CA2.0).² It's a new,

² Machery calls it 'Method of Cases 2.0'.

naturalized form of conceptual analysis. It is easier to say what CA2.0 is not than to say what it is. Here is what CA2.0 is *not*:

- It doesn't aim to deliver a priori analytic truths (2017: 209),
- It's not a semantic notion: it doesn't aim at giving analytic or conceptual truths
- It's not an epistemological notion: it doesn't aim at giving a priori known/justifiedly believed propositions

Instead, what it does is provide us with 'empirical truths about the mind'. Concepts, as Machery construes them, are

retrieved by default from long term memory to play a role in cognition and language understanding ... A body of information is retrieved by default if and only if retrieval possesses the three following properties:

- Speed: Default information is quickly retrieved from long term memory .
- Automaticity: Default information is automatically retrieved from long term memory
- Context independence: Default information is retrieved from long term memory in every context.(Machery 2017:210-11)

CA2.0 has both a *descriptive* and a *revisionary* component: we can describe the content of the information that constitute a concept, but, just as importantly, we can also be prescriptive, or revisionary. In the revisionary phase of CA2.0, we can identify some beliefs that we *should* adopt when we think about the things that fall under the concept.

The Disconnect Between CA2.0 and the Revisionist Tradition

In chapter 7 it is important for Machery to align the revisionary component of CA2.0 with a revisionary tradition in 20th Century philosophy, stretching from Carnap on explication up to Haslanger and other contemporary revisionists. Machery aims to show that this tradition can be naturalized by focusing on the amelioration of concepts construed as described above (i.e., as a body of information retrieved by default, where that is characterized by speed, automaticity, and context independence.)

I share Machery's enthusiasm for the revisionary tradition. Following Blackburn (1999), Brandom (2001) and Scharp (2013), I call that project *conceptual engineering* (see Cappelen 2018). There is, however, an important disconnect between the project Machery describes (revising the things that he calls 'concepts') and the revisionary tradition that he appeals to and relies on throughout that last chapter. The conceptual engineering tradition is focused on concepts as determinants of extensions and intensions. It is essential to Carnapian explication that the semantic value changes (for example, there can be no precisification without such change.) (see Cappelen 2018:11 for Carnap on explication.) It is essential to Sally Haslanger's ameliorative proposals for gender and race terms that their intensions (and probably also extensions) change – that's why she can describe her aim as that of getting rid of women (see Haslanger 2000). The same can be said about all the various cases in that tradition (for more detail see Chapter 2 of Cappelen 2018). It is also true of the other case Machery discusses extensively: causation. One of his examples is Ned Hall's claim that 'there is more than one concept of causation' (2017:222). But when Hall says that, he is not talking just about fast, automatic, and context-insensitive beliefs about causation: it is crucial that the two concepts of causation have different extensions and intensions. In sum: the project Machery advocates isn't the project any of these revisionists are engaged in.

One additional point about this: If the revisionary tradition in philosophy, as I just suggested, is concerned with revising extensions and intensions, then it might well be the case that revision is not a project in psychology. Those of us who are externalists think that the semantic value of an expression is determined by features that, in Putnam's famous phrase, ain't in the head. These

features could be environmental (Putnam (1975)'s twin earth case) or social (Burge (1979)'s arthritis case) or they could depend on activities over which we have no control, such as baptisms and the causal chains of communication that result from them (Kripke 1980), or again the particular details about how an expression is used by the millions of other speakers of our language (Williamson 1994). If externalism is true, revision is difficult because semantic values are determined by something outside our control, and so, pace Machery, revision is not a project in psychology.³

The Disconnect Between CA2.0 and Philosophical Practice

Much of Chapter 7 is devoted to an account of how CA2.0 can help us do philosophy. According to Machery CA2.0 helps make our concepts explicit and this can have a therapeutic effect because it will help dispel philosophical puzzles. It might, as an illustration, reveal that we have several distinct concepts of causation and this explains ‘why we seem committed to incompatible claims about the properties of causation as well as why straightforward philosophical accounts of causation appears prone to counterexamples, resulting in added epicycles and convolutions’ (2017: 222). Machery also recommends CA2.0 because it helps us assess the inferences concepts dispose us to draw, which will help us modify the inferences that are, in some way, defective. As an illustration he uses inferences we make involving the concept of ‘race’ and ‘consciousness’.

I have my doubts about these alleged applications of CA2.0. I think the use of ‘concept’ gives a veneer of plausibility to the claims Machery makes on behalf of CA2.0. If we constantly remind ourselves that what Machery means by this: a subset of judgements that are made fast, by default, and in a context-insensitive way, the claims he makes start looking bizarre. This is because most philosophers think slowly, carefully and endlessly assess the judgements we come up with. They are not what Machery calls ‘default’ judgements. Of course philosophers are normal people who in their day-to-day life will judge fast and rely on defaults. However, take the claims made in this review: it took a really long time to write, I’ve thought about every sentences

³ If this went by a bit fast, see Part II of Cappelen 2018 for elaboration.

in here for a long time, I've read the whole thing over many times, I've had other people read it, etc. what I just said is of course entirely trivial: we all know that these are features of the claims made in theoretical work by philosophers. We don't rely on fast and default judgments.

Philosophers who theorize about race or consciousness or any of the topics that CA2.0 allegedly should help us with think carefully and slowly. They rely on arguments that are assessed carefully over long periods of time. What matters is not the default (fast, automatic) judgment we make. What matters to us (or should matter to us) is the truth of what we say, and the validity and soundness of our arguments. CA2.0 won't help determine whether claims are true or arguments valid. It can help us assess the truth of claims about what people's default judgements are, but that's not what philosophical theories of, for example, race or consciousness are about.

My underlying worry about the last chapter can be put like this: Machery wants to destroy our cake and eat it too. On the one hand he argues that what we philosophers care about is pointless because modally immodest and so impossible (that's the chapter 6 message). On the other hand, he also wants to say that CA2.0 can help us with the questions we care about and work on (causation, consciousness, etc). The problem, I've argued, is that CA2.0 doesn't help answer those questions. Chapter 7 is in large part an effort to do something Machery shouldn't aim for given the conclusion in chapter 6. If he wants to burn down our house, the last chapter better not be a proposal for how to decorate it. He should just build a new house, in a different place.

Bibliography

- Blackburn, Simon (1999). *Think: A Compelling Introduction to Philosophy*. Oxford University Press.
- Braman, Robert B. (2001). Modality, normativity, and intentionality. *Philosophy and Phenomenological Research* 63 (3):611-23.
- Burge, Tyler. (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4 (1):73–122.
- Cappelen, Herman (2012). *Philosophy Without Intuitions*. Oxford University Press UK.
- Cappelen, Herman (2018). *Fixing Language: An Essay on Conceptual Engineering*. Oxford University Press UK.
- Cappelen, Herman & Deutsch, Max Emil (2018). The Crisis of Method in Contemporary Analytic Philosophy. *Notre Dame Philosophical Reviews* 1 (05).
- Deutsch, Max (2015). *The Myth of the Intuitive*. The MIT Press.
- Haslanger, Sally (2000). Gender and race: (What) are they? (What) do we want them to be? *Nous* 34 (1):31–55.
- Knobe, J. & Nichols, S. (2008). An Experimental Philosophy Manifesto. In Knobe, J. & Nichols, S. (ed.) *Experimental Philosophy*. New York: Oxford University Press, p. 3-14.
- Kripke, Saul A. (1980). *Naming and Necessity*. Harvard University Press.
- Machery, Edouard ; Mallon, Ron ; Nichols, Shaun & Stich, Stephen (2004). Semantics, cross-cultural style. *Cognition* 92 (3):1-12.
- Machery, Edouard (2017). *Philosophy Within its Proper Bounds*. Oxford University Press.
- Nagel, Jennifer (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research* 85 (3):495-527.
- Putnam, Hilary (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science* 7: 131–93.
- Scharp, Kevin (2013). *Replacing Truth*. Oxford University Press UK. Oxford: Oxford University Press.
- Williamson, Timothy (1994). *Vagueness*. Routledge.