Responses to Herman Cappelen and Jennifer Nado

Edouard Machery¹

Accepted: 12 October 2020 / Published online: 30 October 2020
© Springer Nature B.V. 2020

I am grateful to Hermann Cappelen and Jennifer Nado’s insightful and challenging commentaries. It is difficult to imagine a better pair of symposiasts: a forceful, but perceptive critic of experimental philosophy (Cappelen) and a friendly fellow traveler with an original take on experimental philosophy and its metaphilosophical significance (Nado). In this article, I respond to their criticisms. I first discuss how to delineate the target of the critical arguments inspired by experimental philosophy. I then examine what I have called the disturbing characteristics of experimental philosophy. In Sect. 3, I examine whether my arguments lead me to an untenably broad skepticism. In Sect. 4, I discuss whether the metaphilosophical debate should be framed in terms of a new concept, the concept of knowledge*. Finally, I compare a semantic and a psychological approach to conceptual engineering.

1 The target of the experimental-philosophy arguments

Since the very early days of experimental philosophy proponents of the so-called negative program in experimental philosophy have claimed that their experimental findings—the influence of demographic variables and of trivial changes to the cases (which I call, respectively, “demographic and presentation effects”) on the judgments elicited by philosophical cases (henceforth, “case judgments”)—raise troubles for the argumentative use of cases in particular areas of philosophy (epistemology: e.g., Weinberg et al. 2001; philosophy of language: e.g., Machery et al. 2004) or more generally in philosophy (e.g., Weinberg 2007; Machery 2011). It has however always been difficult for this philosophical tradition to determine precisely and in a principled manner what class of judgments was supposed to be

✉ Edouard Machery
machery@pitt.edu

¹ University of Pittsburgh, Pittsburgh, USA

undermined by experimental philosophers' demographic and presentation effects, and some have used this difficulty against the negative program, arguing that any argument based on demographic and presentation effects would impugn judgment in general (Williamson 2007).

The difficulty under consideration does not result from idiosyncratic features of the arguments against the method of cases; it is rather an instance of a general problem related to induction: specifying the reference class for which some generalization is meant to hold on the basis of a subset of this class (e.g., Hempel 1965). For this reason, I will call this problem "the reference class problem." In some scientific contexts, the reference class problem is solved by formal sampling (as noted by Alexander and Weinberg, forthcoming): either by creating a sufficiently large random sample or by some form of matched sampling from some predetermined population. But this solution is not available here since the selection of cases examined by experimental philosophers is not the result of formal sampling from a given population.

Philosophy Within Its Proper Bounds (Machery 2017) addresses the reference class problem by identifying the causes of the demographic and presentation effects—the disturbing characteristics—and by generalizing to all the cases that have the same characteristics. I take this argumentative strategy to be one of the original ideas of the book, and I am delighted to see Nado agreeing. For short, I often refer in the book and elsewhere to philosophical cases generically and I claim to target the method of cases, but, as is the case of generics in general, the claim that philosophical cases elicit unreliable judgments allows for exceptions. Indeed, I repeatedly exclude from the skeptical conclusions of my three main arguments (Unreliability, Dogmatism, and Parochialism) the cases that are known not to elicit demographic and presentation effects (e.g., p. 92, p. 127, p. 140), and I call them "good cases" (p. 120). The Gettier case may be a good case (Machery et al. 2017a, b; but see Machery et al. 2018; for a recent discussion, see Hawke and Shoenen, in press).

In her penetrating commentary, Nado expresses doubts about whether it is possible to treat the class of judgments elicited by cases with disturbing characteristics (which, following my habit, I will simply call "philosophical cases") uniformly and thus to accept the same skeptical attitude toward all case judgments. As Nado puts it, "Suppose I come to have a view on a 'disturbing' philosophical case based on an inference from further theoretical commitments, or on some other chain of argumentation. Are such judgments to be dismissed along with the more intuitive snap judgments that x-phi originally aimed to target?" She illustrates this concern by referring to a judgment about the Gettier case formulated on the basis of theoretical virtues like simplicity: Gettier cases describe an instance of knowledge because the identification of knowledge with true justified belief is much more general and simpler than any other proposal denying knowledge in Gettier cases (Weatherson 2003). The question, then, is whether this judgment is included in the class of judgments impugned by Unreliability, and why.

I will assume that Nado does not take Weatherson's proposal to be a normative proposal about how to use "knowledge" (i.e., it is not a piece of conceptual engineering or a metalinguistic negotiation) nor is it a case of precisifying an

antecedently indeterminate question; it is rather a truth-evaluable claim about the fact of the matter in the situation described in Gettier cases, and it takes theoretical virtues to be truth-conducive. This judgment and similar ones are clearly a challenge for the approach proposed in *Philosophy Within Its Proper Bounds*. At least initially, it is not clear whether my skeptical conclusions extend to them, and if they do not, on what grounds the arguments I proposed can be properly circumscribed. I will have more to say about this question below, but as a preamble it is important to state that it is wrong to focus on these unclear cases (see my discussion of Williamson's points in Machery 2017, 183): whether or not we know where to draw the border of the reference class is irrelevant for how we should treat prototypical cases.

Still, what about Weatherson's case judgment and its ilk? On second thought it should clearly be excluded from the class of case judgments threatened by the arguments of Unreliability, Dogmatism, and Parochialism, for the process by which Weatherson forms a judgment about whether the character in a Gettier case genuinely has knowledge differs markedly from the process by which we ordinarily ascribe or deny propositional knowledge. Indeed, Weatherson must override, so to speak, the judgment he would otherwise make to conclude that the agent does know the relevant proposition. It would be a mistake to say that Weatherson reaches his conclusion by a process that is altogether different from the process underlying everyday judgments about knowledge: after all, he must take into account various cues about the epistemic situation of the agent in a Gettier case, as we do when we ascribe and deny knowledge in everyday situations and when we make a judgment about the Gettier case, and he must have started with, or at least considered or relied on at some point, the ordinary understanding of knowledge in order to modify it. But the conclusion is ultimately driven by a theory about knowledge reached on the basis of, among other things, theoretical virtues. The ordinary understanding of knowledge features in a different manner from when we decide on the basis of a *New York Times* article that Trump does not know what he is talking about when he blabbers about COVID-19. And this judgment is thus more similar to the kind of judgment Lewis called "spoils to the victor" (Lewis 1986, 194): Once we have a theory, reached on some ground or other (by considering cases and using theory-construction tools like theoretical virtues), we can apply it to unclear cases.

Additionally, while the skeptical conclusions reached as a result of Unreliability, Dogmatism, and Parochialism do not apply to the theory-driven case judgments, I have also argued that theoretical virtues cannot be in general assumed to be truth-conducive (Machery 2017, Chapter 6). If they are, it is only locally, for reasons that can be explained, and there is no reason whatsoever to think that they are generally truth-conducive in philosophy.

While Nado might agree with this diagnostic, she is likely to be unsatisfied until a more general characterization of the reference class to which the skeptical conclusions of Unreliability, Dogmatism, and Parochialism are meant to apply. I am not confident I can settle all possibilities in advance, but I can at least exclude the following class of case judgments: judgments that are formed on the basis of an explicitly developed theory that differs from philosophers' antecedent understanding of the theory's topic. Such case judgments have to be assessed on other grounds,

perhaps because they have been reached by means of other case judgments and theoretical virtues (as is the case for Lewis).

This discussion, however, addresses only part of Nado's broader concerns: reliability is likely to vary among case judgments, even among those that are not theory-driven. As she puts it (emphasis in the original): "[W]e have *prima facie* reason to assume that *some* such cognitive process types will substantially differ in their vulnerability to the disturbing characteristics" (see also Alexander and Weinberg, forthcoming for a similar concern). It would not bet against the existence of some differences in reliability among partitions of the class of case judgments, even at good odds, but there are differences and differences. First, some differences in reliability are probably too small to be of interest: we can group partitions into equivalence classes based on the size of the variation in reliability. Second, Nado is fully aware that differences would only matter if case judgments in some partitions were "sufficiently reliable to permit their use," and that we have no evidence at this point that there are such partitions.

In his insightful commentary, Cappelen too raises concerns about the homogeneity of the cases used in philosophy. He doubts that they have any characteristic in common, and as a result that they can impugned as a class: "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 case' doesn't describe 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." Nado and Cappelen are right to press this point: this is a crucial issue, one that would weaken experimental philosophers' critic of the methods of analytic philosophy. However, even if experimental philosophers' generalization failed, their criticisms would not be fully invalidated: they would still apply to the cases they have focused on, which happen to be central to important literatures in philosophy, such as the Gettier case, the Gödel case, or the Trolley cases, and to the probably many other cases that are like them (Machery 2017, 181). In any case, I have tried to meet this concern halfway. I distinguish various uses of cases in philosophy in Chapter 1 of *Philosophy Within Its Proper Bounds*, many of which I deem to be uncontroversial: for instance, using cases to elicit puzzlement or to illustrate a definition or a theory. What's more, some famous cases in philosophy such as Davidson's original swampman case are just used in such uncontroversial ways. So, I limit my concerns to the use of cases for argumentative purposes. The discussion of Nado's objection has also led me to exclude another type of case judgment. Finally I grant that there might be good cases, and I only generalize to those cases that have some of the disturbing characteristics—what philosophers should stop using are *those* disturbing cases—although I believe philosophers are extremely likely to rely, for non-accidental reasons, on such cases (more on this in the next section). I suspect that Cappelen will not be willing to meet me halfway, insisting either that the disturbing characteristics I have identified do not undermine the reliability of case judgments or that many cases in philosophy can be used argumentatively without having any disturbing characteristic. I turn to these two issues in the next section.

2 The disturbing characteristics

As noted in Sect. 1, I call “disturbing characteristics” the properties of cases that explain why the philosophical cases examined by experimental philosophers elicit demographic and presentation effects. I acknowledge in Chapter 3 of *Philosophy Within Its Proper Bounds* that little empirical evidence bears on what these disturbing characteristics are, but I speculate that three features of the philosophical cases examined by experimental philosophers are disturbing:

- *Unusualness*: the cases used by philosophers describe unusual situations.
- *Separating what goes together in everyday life*: the properties that are possessed by instances of a concept in everyday concept applications are separated in philosophical cases.
- *The entanglement of target and superficial content*: the features of cases that are relevant to philosophical cases and the philosophically irrelevant narrative settings of these cases both influence judgment.

I am explicit that not all philosophical cases possess these disturbing characteristics: “some philosophical cases may well be free of all these characteristics, and the judgments these elicit are not vulnerable to the influence of demographic and presentation variables” (2017, 112). Those that don’t are likely to be good cases, but I argue that to fulfill some of their argumentative functions cases must have some of these disturbing characteristics (others are accidental): “some disturbing characteristics are accidental, and cases without them could be developed; others are non-accidental: to bear on some central material-mode issues, cases must have these properties” (2017, 112).¹ I also explicitly reject the idea that these three disturbing effects are always found in all the cases known to give rise to demographic and presentation effects: “some philosophical cases may possess all these characteristics, while others only a few” (2017, 112). I do not exclude the existence of other disturbing characteristics. Finally, the disturbing characteristics do not necessitate unreliability; they just increase the likelihood that judgments are going to be unreliable (2017, 112): “these characteristics do not necessitate that demographic and presentation variables influence judgment rather, they merely make it likely.”

Cappelen takes issues with various aspects of the discussion of the disturbing characteristics, which he rightly identifies as one of the pivots of the book. He focuses on unusualness, claiming that the points generalize to the two other disturbing characteristics: “what I will say will generalize, and if so, it has wide reaching implications for the overall project in the book.” His focus is

¹ Cappelen wonders whether this modal claim is consistent with the claim in Chapter 6 that many modal propositions are beyond our epistemic reach. The answer is, Clearly yes: the modal skepticism is limited to *metaphysical necessities* (and not even to all of them), while the necessity expressed by the claim that cases must have some disturbing properties isn’t. I need not make any claim about what holds in every possible situation to claim that for us philosophers cases must have these characteristics to fulfill their argumentative functions.

understandable since I discuss Cappelen's (2012) take on the unusualness of philosophical cases, but he does not explain how his objections would generalize to the two other disturbing characteristics, and it is not straightforward how they would since they depend on some specific points of my discussion of unusualness. So, even if Cappelen were right about unusualness, more would need to be said to defang the discussion of the disturbing characteristics in Chapter 3.

More important, Cappelen loses track of the dialectical situation. The disturbing characteristics are hypothesized to explain the unreliability of the judgments elicited by the philosophical cases examined by experimental philosophers. Demographic and presentation effects are meant to establish that these philosophical cases elicit unreliable judgments. Even if Cappelen were right about unusualness and even if his arguments somehow extended to the two other disturbing characteristics, this would only show that I have failed at identifying the features of the cases examined by experimental philosophers that produce unreliable judgments, not that there are no such features. Indeed, there must be some if I am right about demographic and presentation effects, something that Cappelen does not discuss. Ultimately, then, it does not matter much if Cappelen is right about unusualness: the disturbing characteristics, whatever they are, explain why the judgments elicited by the philosophical cases examined by philosophers are unreliable, and we should suspend judgment about any case that possesses them.

Perhaps this response is too quick: Cappelen might respond that how threatening the disturbing characteristics are depends on what they are. If they are idiosyncratic features of the cases examined by experimental philosophers, then the threat for the method of cases is limited, although not null as we saw in the previous section. Some cases would be problematic—the ones examined by experimental philosophers and perhaps a few others—not the bulk of philosophical cases. What makes the three candidate disturbing characteristics threatening is that they are plausibly shared by many other cases. While this response has merit, its significance is limited: experimental philosophers have examined a limited, but reasonably diverse group of cases, and it would be very surprising, although of course not impossible, if whatever explains why those cases elicit unreliable judgment was only found in them, and not in a much broader group of philosophical cases.

In any case, it is worth looking at Cappelen's two objections to the metaphilosophical significance of unusualness. Cappelen (2012) proposed a counterexample (adding two and three pink elephants) to the claim that unusualness makes judgment unreliable. I responded in *Philosophy Within Its Proper Bounds* that whether a situation is unusual for the purpose at hand depends on what is "germane" to the judgment it is meant to elicit: "the fact that harm involves causing death in the trolley cases is germane to getting it right when one is making a permissibility judgment, while whether one is adding chairs or pink elephants is not when one is adding" (2017, 121). Cappelen responds in two ways: first, he gives another counterexample, where the color and identity of what one is adding seems germane (although he doesn't explain why): "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 judgment, but surely the color of the

elephant is germane.” Second, and more important, he presses me to explain what “germane” means here: how does one determine whether some aspect of a situation is “germane” to the judgment at hand and as a result bears on its usualness: “it’s super important that we’re given an account of germaneness, but Machinery has nothing to say about this.”

Cappelen rightly presses me to give an account of germaneness: in *Philosophy Within Its Proper Bounds*, I failed to make explicit what this means. To address this shortcoming, let’s first introduce the notion of the topical concept: the topical concept is the concept whose application a philosophical case is meant to elicit. The topical concept of a Gettier case is the concept of knowledge; the topical concept of a Trolley case is the concept of permissibility. We can then define the notion of replaceability as follows: an aspect of a case (e.g., the identity of its main character) is replaceable if and only if what a reader views as evidence for or as constitutive of the application of the topical concept is not affected by its replacement by something else. Color is replaceable in Cappelen’s two examples: in his first example, the elephants’ color can be replaced with another property (e.g., their size: dwarf elephants) or even eliminated without this change affecting what the reader takes to bear evidentially or constitutively on the result of the addition; in his second example, the elephants’ color can similarly be replaced with another property (e.g., their size: dwarf, normal, and giant elephants). In the first example, color is utterly otiose; in the second example, it distinguishes several sets of elephants, which must be tracked in order to complete the addition, but other properties can distinguish the relevant sets of elephants. (Incidentally, *being an elephant* is also replaceable.) What is not replaceable is information about set membership. In the first example, what is not replaceable is the fact that there is a set of two items as well as a singleton; in the second example, the fact that there is a set of two items as well as a singleton, another singleton, and a set of seven items, and the fact that one of the two singletons and the set of seven items belong to a superset. Change any of this, and what the reader takes to be evidentially and constitutively relevant to the addition changes. I can now give an account of germaneness—the facts that are germane are those that are not replaceable—and distinguish unusual situations from usual ones: in unusual situations, some germane facts are missing or somehow altered. Ned Stark’s lie in the *Game of Thrones* is not unusual because the situation includes all the non-replaceable facts: Ned Starks intends to mislead his wife, Jon, and Robert Baratheon by asserting something (that Jon is *his* son) he believes (indeed, knows) to be false. The teletransporter in *Star Trek* is an unusual vehicle (the topical concept) because some of the facts that readers take to be evidentially or constitutively relevant for judging whether something is a vehicle (e.g., that the candidate vehicle moves) are absent.

Cappelen’s second argument goes as follows: if judgments about unusual situations are unreliable, how can I reliably identify a situation as unusual since that judgment itself should be unreliable by virtue of being about an unusual situation? Briefly, we are reliable at recognizing usual situations as such, and we judge those that aren’t recognized as usual to be unreliable.

3 The costs of skepticism

Bracketing the previous concerns about the boundaries of the reference class discussed in Sect. 1 above, the weight of evidence entitles me to assert that case judgments are unreliable, but it is also the case that new findings could lead us to withdraw this assertion (see, e.g., Knobe, forthcoming; but see Stich and Machery, forthcoming), and some have wondered whether this possibility should lead experimental philosophers to qualify the skeptical conclusions they want to draw (Sosa 2007; discussed in Machery 2017, 171–175). In this spirit, Nado asks: “is it really better to refrain from theorizing entirely rather than run a moderate risk of getting things wrong? Legions of weathermen would likely protest otherwise.”

Nado discusses different types of costs that could result from banning case judgments, ranging from the possibility that this ban might be challenged by new data to its consequences for everyday judgments that seem to share some features with case judgments. I will explore these alleged costs in turn.

When you think it through, the analogy between experimental philosophers and weather scientists does not support Nado’s tolerant use of case judgments. Weather scientists make predictions about the next few days—five to six days ahead in 2020, one or two when I was a kid—but suspend judgment when it comes to the weather two or three weeks ahead. It isn’t that their models are silent about the weather weeks ahead; rather, weather scientists have reasons to believe their models are unreliable when it comes to predicting weather weeks ahead while being reasonably reliable for the days ahead, despite being obviously (and sadly) fallible. As I insist in *Philosophy Within Its Proper Bounds*, the evidence shows that case judgments are unreliable, not just fallible (p. 171–175), and nothing Nado has said challenges this claim. Nado probably thinks that the evidence for the unreliability of case judgments is suggestive, but not compelling, and that weather scientists’ evidence about the unreliability of their predictions is just much (much much perhaps) better than the evidence for the unreliability of case judgments. No doubt weather scientists have more evidence than experimental philosophers (among other things, they can compare their predictions to reality), but their advantage does not mean that experimental philosophers’ evidence is just suggestive. What Nado must establish, then, is that the evidence doesn’t show that judgments are unreliable, but at best suggestive.

It is also important to conceive of the costs of maintaining the status quo in philosophy properly. Opportunity costs should be taken into consideration: instead of worrying merely about the possibility of jettisoning traditional philosophy erroneously and of the small costs of letting it bloom, we should take into account what else philosophers could be doing: research topics selection is a zero sum game, and maintaining the status quo in philosophy in effect means not engaging in more productive ways of doing philosophy (see the discussion in Machery 2017, pp. 190–192; Chapter 7). Perhaps Nado disagrees about the values of those, but she doesn’t say.

What about the other costs that result from suspending judgment whenever there is evidence of disturbing characteristics or whenever there is direct evidence of

unreliability in the guise of presentation and demographic effects? Nado refers to the consequences of suspending judgment for “ordinary folk talk of ethics, of freedom, of knowledge, of beauty, of mental states, of causation”; she notes that “practical matters hang on such variation-sensitive judgments” and “it’s a risk we must take; the short of considered suspension of judgment that Machery recommends isn’t really an option here.”

Nado, who acknowledges making a “self-consciously Williamsonian” point here, is unconvincing for the same reasons that explain why Williamson’s attack on experimental philosophy failed: distinctions must, and can, be drawn, and it isn’t helpful to paint everyday judgment with a broad brush. Some of the assertions she refers to (talk about ethics and beauty) just aren’t expressing judgments at all: they express preferences or perhaps commitments to particular norms. They are not truth-evaluable in any robust sense, and thus are neither reliable nor unreliable. Perhaps we can’t avoid having moral and political opinions that are determined by who we are and a myriad of contingent factors, but this is not an issue for my views.

Naturally, Nado might disagree with this metaethical view about moral and political assertions, and in any case not all assertions she refers to—such as assertions about causality and mental states—are plausibly understood as expressions of preferences. So what to do with those? Again distinctions matter. Many judgments about knowledge, causation, and responsibility are very difficult to frame (consider the judgment that stealing is wrong), and to the extent that they vary across demographic groups partitions are often easily justified, and the variation does not justify suspending judgment. Epidemiologists are better placed to judge about the COVID pandemic, and political disagreements about the matter raise no concern for the argument developed here: the judgments of Trumpsters should be ignored, the judgments of epidemiologists taken seriously.

But what about the other lay judgments, those that can be framed or those that vary across demographic groups when it is not clear whether we can partition the class of judgments into one that is reliable and one that isn’t? What’s good for the goose is good for the gander as the saying says: skepticism is not limited to the seminar room, and we should suspend judgment there too. While *Philosophy Within Its Proper Bounds* targets the excess of philosophy, the intellectual humility it recommends in philosophy has a broader significance: we are just too confident in our judgments.

4 Reliability

Nado is concerned about the status of the reliability condition for knowledge, on which Unreliability relies: is it a metaphysically necessary truth? If so, how do I know it given my skepticism about knowing metaphysical necessities? And why would it be a nomologically necessary truth? What is the relevance of the laws of nature for this condition on knowledge?

These are excellent questions, and I was not fully clear about the status of this principle in *Philosophy Within Its Proper Bounds*, but there are a number of satisfying answers to them. First, one could go expressivist about the issue: I am

merely proposing and recommending a particular norm for knowledge. On this view, the claim that reliability is necessary for knowledge is neither metaphysically nor nomologically true because it is no more true than other prescriptions. This proposal coheres with my inclination toward understanding normative discourse in the moral and political domain. A shortcoming of this proposal, however, is that if someone does not share this commitment, Unreliability will be ineffectual. The impact of this shortcoming is somewhat blunted by the fact that Unreliability is only one of the three arguments leading to skepticism about case judgments, and the further fact that the two other arguments do not rely on reliability. Furthermore, most philosophers, internalists as well as externalists, do in fact agree about the significance of either reliability in itself or at least a belief in reliability.

An alternative proposal is inspired by Nado's own commentary: one could view this proposal as a piece of conceptual engineering. A shortcoming, discussed below, is that conceptual engineering requires substantial argumentation for its comparative superiority (see Sect. 5 below), and I did not offer any in *Philosophy Within Its Proper Bounds*.

Finally, one could view this proposal as a piece of non-ideal theorizing: so understood, I am asserting that for us, reliability is necessary for knowledge, while suspending judgment about whether it is for creatures different from us and even about whether this question is meaningful. It might or might not be true in all possible situations that knowledge requires reliability, but that is totally irrelevant for the claim on which Unreliability rests on. As Pogge has described those who resist pressure to look for modally immodest normative principles (2008, 468):

Those (...) are not typically lazily neglecting to unify their moral views about the various factual contexts by tracing them back to ultimate principles that cover all possible worlds. Rather, most of them seem to be (...) reluctant to try to cover, with their moral principles, factual contexts substantially different from our own, like the described outlandish context of regressing fetuses.

On this view, in contrast to the first proposal, it is true that unreliability requires knowledge, and necessarily true for many possible situations, provided the possible worlds are close enough to ours. In contrast to the second proposal, no reform of the concept of knowledge is proposed, and no need to argue for such a defense is needed. This proposal has the advantage and drawback of embracing non-ideal theorizing: if one remains unconvinced by this method and its defense, then one would not accept one of the crucial premises of Unreliability.

5 Knowledge*

Nado makes one of the most interesting proposals in recent metaphilosophy: the debate about the past and prospects of philosophy should not be formulated in traditional epistemological terms (Do we *know* that the agent in the situation described by Russell's clock case fails to know that it's 3:00 pm?), but in newly conceptually engineered terms: Do we *know** that the agent in the situation described by Russell's clock case fails to know that it's 3:00 pm? We could then

stipulate how much evidence, in a context-sensitive manner, would be required to know*.

I am friendly to conceptual engineering or, as I call it in *Philosophy Within Its Proper Bounds*, to prescriptive conceptual analysis (Chapter 7; see next section), and I thus welcome this original idea, which should be investigated further. Full evaluation will however have to wait for the details since the superiority of any conceptual engineering depends on the details of the proposals: typically, there are several ways of reforming a given concept, and the assessment of their comparative merits depends on their exact similarities and differences (see Schupbach 2011 on the notion of explanatory power). Furthermore, the case for the superiority of the concept of knowledge* has to be clear. Otherwise, the metaphilosophical discussion in terms of knowledge* would quickly run into an impasse: the traditional philosopher would concede whatever conclusion we might want to draw in terms of knowledge*, and ask, insistently, whether case judgments amount to knowledge.

6 Conceptual engineering

Cappelen and I independently and at the same time defended the importance of conceptual engineering or as I call it prescriptive conceptual analysis (perhaps evidence that conceptual engineering *really is* important), but we have very different views about what conceptual engineering is about: Cappelen (2018) takes conceptual engineering to be a semantic project and endorses semantic externalism (with the surprising result that conceptual engineering is not about concepts); I take it to be a psychological project focused on the tools we use to reason about the world. To establish the philosophical significance of prescriptive conceptual analysis, I have argued that the project of psychologized conceptual analysis is similar to a long tradition in philosophy, best illustrated by Carnap and Gramsci (Machery 2017, 226): “many traditional philosophical projects are, in important respects, similar to the program described in the present chapter.” I then referred to Hume, Whewell, James, Peirce, and Carnap.

Cappelen raises two main objections. First, psychologized prescriptive conceptual analysis differs from other forms of conceptual engineering in philosophy, including Carnap’s explication, because the latter were explicitly semantic. As he puts it, “The conceptual engineering tradition is focused on concepts as determinants of extensions and intensions. It is essential to Carnapian explication that the semantic values changes (for example, there can be no precisification without such change).” Second, because of the notion of concept I rely on, the project of psychologized prescriptive conceptual analysis seems at odds with what philosophers do. Following Machery (2009), I identify concepts as bodies of information that are retrieved by default from long-term memory, to engage in various cognitive tasks (reasoning, categorization, etc.) where a body of information is retrieved by default if it is retrieved in a context-independent manner, and as a result quickly and automatically. Cappelen objects that so understood concepts don’t seem to be what philosophical reflection with its slow, careful, controlled pace is about: “if we constantly remind ourselves that what Machery means by this

[“concept”]: a subset of judgments that are made fast, by default, and in context-insensitive way, the claims he makes start looking bizarre. This is because most philosophers think slowly, carefully and endlessly assess the judgments we come up with. They are not what Machery calls ‘default’ judgments.”

Cappelen is right that my psychological, explicitly non-semantic take on conceptual engineering differs from some older and contemporary versions of conceptual engineering. Similarity is however always a matter of respect, and while different from the latter in some respects, it is similar in other respects. One of the goals of conceptual engineering in philosophy and in the sciences is to avoid being caught in paradoxes and to avoid unreliable inferences, and my psychological approach is tailored to meet these goals: by engineering a concept, we determine what assertions people make and prevent them from asserting something and its negation; we also determine what conclusions they draw from which premises (see the example of the concept of innateness in Chapter 7 of *Philosophy Within Its Proper Bounds*), ensuring that their inferences are tuned to the world we live in, as revealed by our best science, or the world we want to live in, as determined by our political preferences.

What’s more, we can do without the semantic detour: we can reformulate the outcomes of conceptual engineering that appear to require a semantic approach. Consider precisification: instead of saying that a concept is engineered so as to turn indeterminate propositions into propositions that are determinately true or false, we now say that our goal is to be able to make determinate judgments (judgments we take to be determinately true or false) where we were unable to judge determinately beforehand (because we took the matter to be indeterminate). Consider also a change in the truth value of propositions: instead of engineering the concept of a woman so as to make it true that there would be no women if sexist subordination disappeared, we engineered the concept of a woman such that if we were to believe that sexist subordination has disappeared, we would be disposed to believe and assert that there are no women. Semantic outcomes are replaced by changes in our dispositions to infer, believe, or assert. Comparing these two approaches is beyond the scope of this symposium, but I speculate that this psychologized reformulation of the semantic outcomes of traditional conceptual engineering gets us everything we should want. Consider the precisification of the concept of a heap such that nine grains together do not make a heap but ten or more do. What does saying that “these nine grains are not a heap” has been rendered determinately true add to saying that we are now likely to ascent to “these nine grains are not a heap” as being determinately true? Search me.

Finally, it is worth emphasizing the crucial advantage of a non-semantic approach: the conceptual engineer proposes to modify the conceptual role of a concept without having to worry whether this conceptual role is constitutive of the meaning of the concept. Unending controversies about what constitutes the meaning of a concept are entirely bypassed, and we can focus on what matters: avoiding contradictions, inferring reliably, etc.

Cappelen’s second argument fails in the context of conceptual engineering. (It would be a better argument in the context of descriptive conceptual analysis). There is no inconsistency, not even a tension, in holding that, on the one hand, the targets

of conceptual engineering are bodies of information retrieved by default and that, on the other, the process of conceptual engineering requires much pondering: we need to decide and argue which revision of our practices is the best.

What seems to underlie Cappelen's second argument is the idea that it is strange for philosophers to care about concepts, as I characterize them: he speaks dismissively of "fast and default judgments." This concern can be alleviated by seeing philosophy as concerned, at least in part, with describing and improving the cognitive tools people, not simply philosophers, use when they engage with the world. These tools should be concepts as I understand them: they should drive people's attention to the world, their understanding of what they perceive, their reactions and expectations. For instance, when Gramsci proposed to conceptually engineer the concept of ideology, his goal was to offer a way of describing discourses about social reality (as "ideologies") that would give a task to political activists (viz. identify and criticize ideologies). This is also true of the concepts of consciousness that, following Machery (2017, 215–216), Cappelen discusses: Block's distinction between phenomenal and access consciousness is meant to be a tool to prevent fallacies not only in philosophy, but also in neuroscience. It is meant to become a default way of understanding phenomena related to consciousness. It is only a useful conceptual distinction if the mechanisms that underlie the qualitative aspects of perceptions differ from those that underlie the use of information in action and reasoning.

References

- Alexander, J., & Weinberg, J. (in press). You just can't count on (un)reliability. *Analysis*.
- Cappelen, H. (2012). *Philosophy without intuitions*. Oxford: Oxford University Press.
- Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford: Oxford University Press.
- Hawke, P., & Schoonen, T. (in press). Are Gettier cases disturbing?. *Philosophical Studies*.
- Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: The Free Press.
- Knobe, J. (forthcoming). Difference and robustness in the patterns of philosophical intuition across demographic groups. *Review of Philosophy and Psychology*.
- Lewis, D. K. (1986). Causation. In D. K. Lewis (ed.), *Philosophical papers*, vol. 2 (pp. 159–213). Oxford: Oxford University Press.
- Machery, E. (2009). *Doing without concepts*. New York: Oxford University Press.
- Machery, E. (2011). Thought experiments and philosophical knowledge. *Metaphilosophy*, 42(3), 191–214.
- Machery, E. (2017). *Philosophy within its proper bounds*. Oxford: Oxford University Press.
- Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition*, 92, B1–12.
- Machery, E., et al. (2017a). Gettier across cultures. *Nous*, 51(3), 645–664.
- Machery, E., et al. (2017b). The gettier intuition from South America to Asia. *Journal of Indian Council of Philosophical Research*, 34(3), 517–541.
- Pogge, T. (2008). Cohen to the rescue! *Ratio*, 21, 454–475.
- Schupbach, J. N. (2011). Comparing probabilistic measures of explanatory power. *Philosophy of Science*, 78(5), 813–829.
- Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies*, 132, 99–107.

- Stich, S. P., & Machery, E. (forthcoming). Demographic differences in philosophical intuition: A reply to Joshua Knobe. review of philosophy and psychology.
- Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies*, 115(1), 1–31.
- Weinberg, J. M. (2007). How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy*, 31, 318–343.
- Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, 29(1/2), 429–460.
- Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.