Why the method of cases doesn't work

Suhler, Christopher Louis

2019

Suhler, C. L. (2019). Why the method of cases doesn't work. *Review of Philosophy and Psychology*, 10(4), 825-847. doi:10.1007/s13164-019-00433-0

<https://hdl.handle.net/10356/142430>

<https://doi.org/10.1007/s13164-019-00433-0>

© 2019 Springer Nature B.V. This is a post-peer-review, pre-copyedit version of an article published in *Review of Philosophy and Psychology*. The final authenticated version is available online at: <http://dx.doi.org/10.1007/s13164-019-00433-0>

Downloaded on 26 Apr 2023 19:12:59 SGT

Why the Method of Cases Doesn't Work

Christopher Suhler

Philosophy Programme
School of Humanities
Nanyang Technological University
14 Nanyang Drive, Singapore 637332
csuhler@ntu.edu.sg

ORCID: 0000-0001-7425-0668

Abstract: In recent years, there has been increasing discussion of whether philosophy actually makes progress. This discussion has been prompted, in no small part, by the depth and persistence of disagreement among philosophers on virtually every major theoretical issue in the field. In this paper, I examine the role that the Method of Cases (MoC) – the widespread philosophical method of testing and revising theories by comparing their verdicts against our intuitions in particular cases – plays in creating and sustaining theoretical disagreements in philosophy. Drawing on work from psychology, I argue that there is a fundamental incompatibility between (a) the structure of the theories that philosophers seek to construct using the MoC and (b) the structure of the concepts on which our case-specific intuitions are based. This incompatibility renders MoC-based philosophical theorizing unable ever to succeed by the very standards of adequacy that it sets for itself. And this, in turn, helps to explain the depth and persistence of theoretical disagreements – and, in certain ways, the lack of progress – in the many areas of philosophy where the MoC plays an important role.

Keywords: method of cases, intuitions, concepts, counterexamples, philosophical methodology, philosophical progress, metaphilosophy

1. Introduction and Overview

In recent years, there has been increasing discussion of whether philosophy actually makes progress.¹ This discussion has been prompted, in no small part, by the depth and persistence of disagreement among philosophers on virtually every major theoretical² issue in the field.³ Particularly when compared to the sciences – which, by nearly any reckoning, have made and continue to make substantial progress – philosophy enjoys far less consensus on fundamental theoretical issues and has a track record of far fewer “settled” questions (Chalmers 2015, Daly 2017). While consensus is by no means a perfect indicator of a field’s progress, the widespread, persistent theoretical disagreement found in philosophy is, at the very least, something that is worrisome and in need of explanation.⁴

¹ See, e.g., Brennan (2010), Fumerton (2010), Kornblith (2013), Chalmers (2015), Daly (2017), Frances (2017), Shand (2017), and Stoljar (2017). For additional perspectives and issues, see the collection by Blackford & Broderick (2017).

² I will often say ‘theoretical disagreement’ to emphasize that my target is disagreement in philosophers’ answers to theoretical questions such as “what is knowledge?” or “what is causation?”, as opposed to disagreement in case-specific judgments of the sort studied by experimental philosophers. As will be discussed below, a key advantage of my argument is that it doesn’t depend on the existence of the latter type of disagreement (which, for clarity, I will henceforth refer to as ‘variation’ rather than ‘disagreement’).

³ The survey by Bourget & Chalmers (2014) provides a useful quantitative picture of many such theoretical disagreements.

⁴ As Daly (2017) notes, explaining the persistence of disagreement over the answers to major philosophical questions is not just a project for skeptics about philosophical progress. It is a project even – and perhaps especially – for those who seek to defend philosophy as having made (or about to

In this paper, I examine the role that the **Method of Cases** (henceforth ‘MoC’) – the widespread philosophical⁵ method of testing and revising theories by comparing their verdicts against our intuitions in particular cases – plays in creating and sustaining theoretical disagreements in philosophy.⁶ Drawing on work from psychology, I argue that there is a fundamental incompatibility between (a) the structure of the theories that philosophers seek to construct using the MoC and (b) the structure of the concepts on which our case-specific intuitions are based. This incompatibility renders MoC-based philosophical theorizing unable ever to succeed by the very standards of adequacy that it sets for itself. And this, in turn, helps to explain the depth and persistence of theoretical disagreements – and, in certain ways to be described, the lack of progress – in the many areas of philosophy where the MoC is employed.

The claim that case-specific intuitions cannot guide philosophers to the sorts of theories that they seek is, of course, one that has also been advanced by a number of experimental philosophers. These arguments generally proceed by appealing to evidence

make) substantial progress, since it is their position, not that of the skeptics, that seems to be at odds with the field’s track record of unresolved questions (Daly 2017, p. 24).

⁵ For the sake of brevity, I’ll usually just say that the MoC is used by philosophers/philosophy rather than specifying each time that I’m referring to analytic philosophers/philosophy.

⁶ For other recent critical examinations of the MoC, albeit from different perspectives from the present paper, see Baz (2015, 2016a, 2016b, 2017) and Mizrahi (2014, 2015). The use of cases and intuitions has also been the subject of intense criticism from experimental philosophers (e.g., Stich 2013, Machery 2017), although again on different grounds from those that I will appeal to here (see below in main text). Finally, some philosophers (e.g., Cappelen 2012, Williamson 2007, Deutsch 2010, 2015) have argued, primarily in response to the aforementioned challenges from experimental philosophy, that criticisms of the MoC rest on mistaken assumptions about the role that intuitions serve in philosophical argumentation and theorizing (see note 8 below).

for variation in case-specific intuitions between individuals in different demographic groups (e.g., Weinberg et al. 2001, Machery et al. 2004) or presentation conditions (e.g., Swain et al. 2008, Turri & Friedman 2014).⁷ While I'm generally sympathetic to such arguments, the argument I will be making here is not an experimental philosophy-style argument from variation. Rather, my argument aims to show, using psychological work on the internal structure of concepts, that even if such inter- and intra-individual variation does not exist – that is, even if case-specific intuitions are shared/stable in the ways that many experimental philosophers contend that they are not – testing and revising via intuitions *still* cannot lead to theories of the sort that practitioners of the MoC seek. This paper therefore presents a challenge to the MoC that, while compatible with those originating in experimental philosophy, is distinct in the assumptions that it grants to defenders of the MoC and the empirical work on which it relies.

These differences are important, for they allow my argument to avoid certain strategies that are sometimes employed against experimental philosophers' variation-based arguments. First, because my argument does not depend on claims about variation in case-specific intuitions, it is not subject to dispute over how much, or what kind of, variation exists and how this would (or would not) be problematic for philosophers' use of cases/intuitions. Second, because my approach relies on work on conceptual structure that is a matter of longstanding consensus in psychology (see Section 3), there is less room for an

⁷ Although cross-cultural differences are probably experimental philosophy's best-known demographic results, variation in case-specific intuitions has also been found for other demographic categories, such as age and personality type. Presentation condition differences, meanwhile, refer to variation in case-specific intuitions resulting from how a given case is presented to participants, such as order and framing effects. For a systematic review of both types of findings, see Machery (2017, Ch. 2).

opponent to claim that the jury is still out regarding what the relevant empirical research actually shows.⁸

I will proceed as follows. In Section 2, I will describe the MoC and its assumptions, with a particular focus on the interdependent conceptions of theory and testing to which the MoC and its practitioners are committed. I will then, in Section 3, argue that these conceptions of theory and testing presuppose a view of conceptual structure that empirical psychology has roundly rejected. Section 4 will examine the implications of this for the viability of the MoC, arguing that the MoC's assumptions about theory and testing are incompatible with one another in light of the psychological research on concepts discussed in Section 3; this incompatibility helps to explain the persistent theoretical disagreement in MoC-heavy areas of philosophy. In Section 5, I will consider some possible defenses and modifications of the MoC that might be offered in response to the previous sections' criticisms and argue that each does not succeed. Finally, in Section 6, I will suggest that although reducing reliance on the MoC will not be a panacea for theoretical disagreement in philosophy, doing so will be a step in the right direction.

2. Intuitions, Theories, and the MoC

Intuitions⁹ play a central role in much philosophical theorizing. Here, for instance, is how Bealer characterizes intuitions' centrality in many areas of philosophy:

⁸ For an example of how these strategies play out in the debate surrounding experimental philosophy, see the exchange between Nagel (2012, 2013) and Stich (2013).

⁹ Exactly how to define 'intuition' is a topic about which there is limited clarity or agreement in the philosophical literature (for discussion, see, e.g., Kuntz & Kuntz 2011 and Weinberg 2014). For the purposes of this paper's examination of the MoC, I'll simply use 'intuitions' to refer to case-specific

It is safe to say that these intuitions – and conclusions based on them – determine the structure of contemporary debates in epistemology, metaphysics, and philosophy of logic, language, and mind. Clearly, it is our standard justificatory procedure to use intuitions as evidence (or as reasons). (Bealer 2000, pp. 2-3)

More specifically, as Bealer notes, intuitions are widely employed as *evidence* for or against particular philosophical theories.¹⁰ Occasionally, the intuitions in question are general (i.e., not case-specific). More often, however, the intuitions that philosophers use in their theorizing are elicited by particular cases – typically imaginary cases (i.e., thought experiments) designed with the explicit aim of probing philosophical theories. If a theory yields a verdict that matches up with our intuitive verdict about the case, this is evidence for the theory’s adequacy. By contrast, if a theory yields a verdict that conflicts with our intuitive verdict – that is, if the case constitutes a **counterexample** to the theory – this is

judgments to which theoretical verdicts are compared. I hope that this sense of ‘intuition’ is sufficiently clear and familiar to readers in the context of the MoC; however, readers who are particularly averse to intuition talk are free to substitute ‘case-specific judgment’ or another similar formulation, as nothing in my argument hangs on the use of the particular term ‘intuition’.

¹⁰ Over the last decade or so, some philosophers, such as Cappelen (2012), Williamson (2007), and Deutsch (2010, 2015) have challenged the widespread view that philosophers use intuitions as evidence. Responding to these reinterpretations of intuitions’ role would require far more space than I have available here, and in any event such responses can already be found in papers by (e.g.) Climenhaga (2018), Nado (2016), Weinberg (2014), and Chalmers (2014). Thus, for the purposes of this paper, I will simply assume that the standard “intuitions as evidence” view is correct.

evidence against the theory's adequacy.¹¹ In this way, the ability to "capture" our case-specific intuitions serves as a key measure of theories' adequacy across a range of philosophical debates. Pust describes this as follows:

Contemporary analytic philosophy is based upon intuitions. Intuitions serve as the primary evidence in the analysis of knowledge, justified belief, right and wrong action, explanation, rational action, intentionality, consciousness and a host of other properties of philosophical interest. Theories or analyses of the properties in question are attacked and defended largely on the basis of their ability to capture intuitive judgments. (Pust 2001, p. 227)

What makes this approach worthy of being called a full-fledged method, as opposed to a mere tool in the philosophical toolkit, is its iterative nature. When a counterexample to a theory is encountered, the theorist will typically attempt to revise her theory so as to give the intuitively correct verdict in the problematic case. The revised version of the theory is then subjected to a new round of testing against our case-specific intuitions, with further revisions taking place if a counterexample is again encountered. This process can be – and, in many areas of philosophy, is – repeated indefinitely for various competing theories, forming a central part of theory construction and evaluation in these areas.

With these points about intuition, testing, revision, and repetition in hand, we can summarize the steps of the MoC as follows:

¹¹ Some philosophers use the even stronger language of 'refutation' of theories by counterexamples. For examples and discussion, see Mizrahi (2015, § 1).

1. Propose a theory setting out criteria for some phenomenon of philosophical interest (e.g., knowledge is justified true belief).
2. Test whether the theory's criteria yield verdicts that match our intuitions in particular cases, including imaginary cases (i.e., thought experiments) that we devise for the express purpose of testing the theory.
3. If the theory's verdict matches up with our intuitions about a particular case, this is evidence for the theory's adequacy. If the theory's verdict fails to match up with our intuitions about a particular case, this is evidence against the theory's adequacy; if such a counterexample is encountered, one (generally) should revise the theory and repeat the process.

This is just a rough characterization of the steps of the MoC, but it is accurate enough in capturing the core methodology: the positing of philosophical theories setting out criteria for some phenomenon of interest and the testing/revision of those theories through repeated comparison against our case-specific intuitions.¹²

As described by Bealer (2000) and Pust (2001) in the passages above, the MoC shapes debate in many areas of philosophy. This can be seen in the frequent use of cases/intuitions in contemporary philosophical publications, talks, and the like, but it is most easily and typically illustrated by pointing to the fact that some of the most influential arguments in the history of analytic philosophy are paradigmatic MoC arguments. Oft-cited examples

¹² For other concise formulations of the steps of the MoC, see Mizrahi (2014, p. 186) and Baz (2016a, p. 107). (Note that for reasons specific to the project Baz is undertaking in the cited paper, his formulation avoids using the term 'intuition'.)

include Gettier's (1963) argument against the justified true belief theory of knowledge, Frankfurt's (1969) argument against the principle of alternate possibilities, Kripke's (1980) argument for the causal-historical theory of reference, Searle's (1980) argument against computationalism, and Jackson's (1986) argument against physicalism.¹³

Given that the same famous examples tend to be used again and again to illustrate the MoC, a determined skeptic might question whether the seeming prominence of the MoC is just the result of a few extremely influential publications, or, if not that, then at least a relic of a certain period in the field's history where the use of cases/intuitions was ascendant. To this there are a number of responses. The first is simply that philosophy's widespread, ongoing use of cases/intuitions to support and undermine theories is taken by most philosophers writing on the topic to be sufficiently obvious that it does not require extensive, ongoing documentation; one need only flip through a few journals or attend a few conferences to see that the use of cases/intuitions to support and undermine theories is very much alive and well in many areas of contemporary philosophy (e.g., ethics, philosophy of action, epistemology). This sets up the second response, which is that cases/intuitions continue to play a key role not just in philosophical arguments in general, but in *influential* philosophical arguments: more recent examples include Clark & Chalmers's (1998) argument that cognitive processes need not stop at the bounds of the body, Christensen's (2007) argument against justified belief in the face of peer disagreement, and Pereboom's (2001, 2014) argument against the compatibility of free will and determinism, to name just a few.¹⁴ As with the examples from earlier decades, these influential cases/intuitions have

¹³ For further quotes, examples, and discussion of the MoC's use in philosophical theorizing, see Mizrahi (2014, pp. 183-187) and Machery (2017, Ch. 1).

¹⁴ Both Clark & Chalmers (1998) and Christensen (2007) employ various cases throughout the course of their arguments, but their best-known and most-discussed cases are probably the Otto case and the

played a major role in structuring debate on their respective topics. This leads to the third response, which is that although the famous papers standardly cited as examples of the MoC do indeed tend to be from the 1960s through the 1980s, they (and cases on which they rely) have continued to structure debate in the relevant literatures long after their initial publication. For instance, while Gettier's (1963) seminal paper is at this point more than half a century old, *Gettier-style* cases have continued to be a powerful force in philosophical theorizing about knowledge (Weatherson 2003).

Finally, there is a fourth line of response, which is to point to empirical findings on philosophers' use of intuitions over time. In a recent paper, Ashton & Mizrahi (2018), building on the methodology of Andow (2015), examine the use of intuitions in philosophy publications from two major databases (JSTOR and HathiTrust). Their findings stand in stark contrast to the idea that the intuition-centric philosophy is an historical relic; far from peaking in (say) the 1960s, the proportion of philosophy publications containing intuition talk (i.e., use of the word 'intuition' and its cognates) has grown with each successive decade, with the highest proportion being found in the decade from 2000-2009 (the most recent decade studied) and the second-highest in 1990-1999. Still more interestingly, and more important in the context of this paper, is that the same pattern holds true not just for intuition talk in general but more specifically for phrases indicating appeals to intuitions as evidence of the sort characteristic of the MoC ("it seems to me that", etc.): the highest rate is found in 2000-2009 and the second-highest in 1990-1999 (Ashton & Mizrahi 2018, pp. 599-607). As Ashton & Mizrahi (2018, pp. 607-610) readily acknowledge, the results of a single study do not, by themselves, decisively settle debates over how intuitions are used in philosophy. But their results do provide a valuable piece of the puzzle. And they put our

restaurant check case, respectively. As for Pereboom (2001, 2014), his Four-Case Argument, as the name suggests, rests on a sequence of four cases meant to secure his incompatibilist conclusion.

imagined MoC skeptic – who acknowledges that the MoC was methodologically central at the time of famous publications from Gettier (1963), Kripke (1980), and others in the 1960s-1980s but denies that it has continued to be so – in a very difficult spot, since as just discussed intuitions’ prevalence has actually increased rather than decreased in more recent decades.

Why are cases/intuitions so widely used – and so influential in determining the fate of philosophical theories? A key part of the answer lies in the structure of the theories that philosophers aim to construct. Philosophical theories are often definition-like, in that they set out (what the theorist takes to be) the defining features of some phenomenon of philosophical interest (knowledge, causation, cognition, morally right action, moral responsibility, etc.). More precisely, such definition-like theories set out criteria that purport to be necessary and sufficient for something to be an instance of the phenomenon under investigation (e.g., knowledge is justified true belief, morally right action is agent-neutral welfare maximization). Since criteria purporting to be necessary and sufficient must be present in all and only instances of the target phenomenon, a single compelling counterexample can have a profound impact on a theory’s perceived adequacy and on the subsequent course of philosophical debate. If one is able to come up with a case in which something meets the theory’s criteria but (intuitively) is not an instance of the relevant phenomenon, this shows that the criteria are not, in fact, sufficient (e.g., Gettier cases in which an agent has a justified true belief but intuitively does not have knowledge). Likewise, if one is able to come up with a case in which something doesn’t meet the theory’s criteria but (intuitively) is an instance of the relevant phenomenon, this shows that the missing criteria (or criterion) are not, in fact, necessary (e.g., Clark & Chalmers’ cases in which a process is intuitively cognitive despite extending beyond the body). Either way, the theory is shown to be inadequate.

As with the use of famous cases from the 1960s-1980s to illustrate the MoC, a determined skeptic might question whether this characterization of philosophical theorizing as seeking necessary and sufficient criteria still holds or is instead a relic of the past. And indeed, it is somewhat less common to encounter a philosophy publication or talk that explicitly declares “I shall now set out necessary and sufficient criteria for X” than it would have been during, say, conceptual analysis’s 20th century heyday.¹⁵ But analytic philosophy’s drive for theories that combine precision and generality – of a sort that can only be satisfied by necessary and sufficient criteria – remains strong. On the precision side, philosophers want theories that provide clear, precise verdicts about whether something is or is not an instance of a phenomenon of interest: whether a given action is morally right (or not), whether a given mental state is knowledge (or not), whether a given process is cognitive (or not), whether one is morally responsible for a given action (or not), and so forth. On the generality side, philosophers want the collection of things that a theory picks out as (say) morally right actions or cognitive processes to be a unified kind, defined in terms of shared features. And whether explicitly framed as such or not, it is through necessary and sufficient criteria – criteria which are satisfied by all and only instances of the phenomenon under investigation – that theories combine precision and generality of this sort.¹⁶

¹⁵ Though see Sytsma (2010) for discussion of traditional (i.e., non-empirical) conceptual analysis’s recent revival.

¹⁶ What about projects, such as the knowledge first program in epistemology or (many) contemporary virtue theories in normative ethics, that explicitly disavow the ambition to offer (non-circular) necessary and sufficient criteria for some important philosophical phenomenon (in these examples, knowledge and morally right action, respectively)? Part of the answer is simply that the distinctiveness of these approaches only makes sense against the backdrop of a widespread disciplinary ambition to come up with necessary and sufficient criteria for phenomena of interest (see,

The ability to give clear-cut case-specific verdicts is important not just at the level of general theoretical desiderata; it is also crucial to the MoC's methodological viability. As discussed above, the MoC, at its core, involves comparing theories' verdicts against our intuitions; verdicts that match our intuitions are taken as evidence for the theory's adequacy, while verdicts that do not match our intuitions are taken as evidence against the theory's adequacy. Theories that yield clear verdicts are therefore crucial if one is to compare what the theory says to what our intuitions say in a given case and then revise accordingly; without the ability to make unambiguous comparisons between theoretical verdicts and intuitive verdicts, the MoC cannot even get off the ground.¹⁷

As mentioned earlier, once the MoC does get off the ground, it can be applied indefinitely: we don't just stop when a counterexample undermines a theory, but instead revise the theory to avoid the counterexample and then test again (and so forth). The idea is that by repeatedly revising our theories to deal with the counterexamples that they encounter, our theories will slowly but surely improve, with the endpoint of this process ideally being a theory that gives (intuitively) correct verdicts no matter what cases we throw at it. Call this idea – that our theories will, through repeated testing and revision via the

e.g., Williamson's [2000, pp. 2-5] motivation of the knowledge first approach in the opening pages of *Knowledge and Its Limits*). Furthermore, for reasons that will become clear in later sections, my criticisms of the MoC apply to *all* philosophical efforts to provide intuitively adequate necessary and sufficient definitions (see Section 4). So even if something like knowledge can be plausibly argued to escape the expectation of necessary and sufficient analysis in virtue of its conceptual primitiveness, this move will not be available for many of the other phenomena/concepts for which philosophers expect necessary and sufficient definitions.

¹⁷ In normative ethics, for instance, virtue ethical and particularist theories can seem unsatisfyingly imprecise and vague – and their case-specific verdicts often difficult to pin down – when compared to rival theories that are built around necessary and sufficient criteria for morally right action.

MoC, ultimately converge on a theory that is fully intuitively adequate (or at least very close to it)¹⁸ – the **Convergence Assumption**.¹⁹

Framed in this way, the use of thought experiments and intuitions to repeatedly test and improve philosophical theories may be thought to parallel the use of laboratory experiments and observations to repeatedly test and improve scientific hypotheses. But despite its superficial similarity to the scientific method, the MoC has a rather less inspiring track record. In a wide range of scientific fields, revising theories to better account for experimental observations has led to clear progress and to important theoretical questions

¹⁸ One may reasonably worry that fully capturing our intuitions is too high a standard and does not reflect what philosophers actually expect of theories with regard to intuitive adequacy. Complicating matters further, there may not even be a single standard of intuitive adequacy across all the areas of philosophy that employ the MoC: it may be that, as Weatherson (2003, pp. 1-2) suggests, full intuitive adequacy is commonly expected of theories in areas such as epistemology and metaphysics, while in areas such as ethics the standard of intuitive adequacy is somewhat less stringent. However, such complications do not affect my main argument in this paper, since as I'll discuss in Section 5, the MoC's problems are not resolved by modifying our standard of intuitive adequacy.

¹⁹ An especially vivid and self-contained example of the Convergence Assumption at work can be found in Parfit's (2011) *On What Matters* (Vol. 1), in which numerous thought experimental cases are used to refine versions of consequentialism, deontology, and contractarianism until they (at least by Parfit's contention) converge onto unified set of necessary and sufficient criteria for morally right action (the "Triple Theory"). Another clear example of the Convergence Assumption as an organizing principle of philosophical inquiry is the post-Gettier efforts of epistemologists to come up with an analysis of knowledge that is immune to counterexamples, which has involved countless cycles of theory, counterexample, and revision. As a final example – this time within a particular theoretical tradition – consider the ever-growing menagerie of consequentialist theories of moral rightness, whose emergence has been driven largely by the susceptibility of previous formulations of consequentialism to counterexamples.

being settled to the general satisfaction of the relevant research communities. By contrast, in philosophy, revising theories to better account for our case-specific intuitions has not led to clear progress or to the settling of important theoretical questions.²⁰

The failure of the MoC to lead to clear progress and settled questions can be traced, in significant part, to the fact that no theory ever clears the bar of intuitive adequacy that the MoC sets out. No matter how plausible a proposed theory may seem, or how many times it has been revised to deal with previous counterexamples, philosophers are always able to come up with further counterexamples in which the theory's verdicts clash with our intuitions. This leads to further rounds of revision to which more counterexamples are then devised, and so forth *ad infinitum*. (Call this the **Problem of Counterexamples**.) The result is stalemates in numerous areas of philosophy between theories that are, at least by the MoC's standard of intuitive adequacy, flawed.

3. The MoC and the Psychology of Concepts

The inevitability with which theories succumb to counterexamples would seem to provide strong inductive reason for pessimism that any philosophical theory will *ever* be deemed adequate by the standards of the MoC. Yet despite this, the MoC continues to be widely used in many areas of philosophy. It is difficult to make sense of this unless philosophers employing the MoC believe that its track record is not, in fact, an accurate indication of its future potential for success. In other words, philosophers who use the MoC

²⁰ See Chalmers (2015) and Daly (2017) for further discussion of unflattering comparisons between scientific progress/consensus and philosophical progress/consensus. For a more optimistic perspective, see Stoljar (2017).

seem to be proceeding on something like the Convergence Assumption: even if none of our theories are currently adequate by the lights of the MoC, we are, through repeated cycles of counterexamples and revisions, working our way toward theories that are. In what follows, I want to argue that the Convergence Assumption is in fact false, and that any hope to the contrary simply reflects misguided assumptions about the concepts with which the MoC interacts.

Why the shift from talk of phenomena (in the previous section) to talk of concepts? The reason is that even if our theories are ultimately aiming to provide accounts of philosophically interesting phenomena (knowledge, morally right action, causation, moral responsibility, etc.), what the MoC is actually tracking, via its use of intuitions, is not those phenomena themselves but rather our concepts of those phenomena. Baz puts this as follows, in response to philosophers such as Cappelen (2012) and Williamson (2007) who claim that the MoC does not investigate our concepts:

I think it is fair to expect of someone who wishes to dismiss the widespread view that the method of cases centrally aims at analyzing or elucidating our concepts—a view to which Jackson has referred with right as ‘the conventional wisdom’ (2009, 105)—to do more than insist that philosophers have mostly and primarily been investigating *knowledge, meaning, freedom, causation, and so on*, as contrasted with *our concepts* of these “objects”. More than two hundred years after Kant, the tendency among philosophers to imagine themselves investigating things as they are in themselves—as opposed to phenomena that are shaped by human sensitivities, capacities, needs, and interests—is surely no evidence that this is what they actually are doing, or indeed that there is anything clear to be done that fits that self-image. (Baz 2016a, p. 110, emphasis and citation in original)

In this passage, Baz frames the MoC's relationship to concepts in terms of the nature and ambitions of philosophical inquiry itself. At root, however, he is making a point about concepts and language that, in any context except for heated debates over how to characterize philosophical methodology, would be so obvious and uncontroversial as to not warrant mentioning. The core point is as follows. We make categorization judgments on a near-constant basis – this is an apple, that is a car, etc. – and these judgments are uncontroversially taken, by psychologists as well as philosophers, to reflect the concepts that we possess (APPLE, CAR, etc.).²¹ This same process of categorization is at the heart of the MoC, except that the categorization judgments that we are making in philosophical cases – X is an instance of morally right action, Y is not an instance of cognition, etc. – reflect concepts of philosophical interest (MORALLY RIGHT ACTION, COGNITION, etc.) and are often dressed up with the label 'intuitions'. And if the MoC is devising theories that attempt to capture these categorization judgments, it is aiming at elucidating our concepts – whether philosophers employing the MoC in their theorizing think of it as such or not.²²

If, for the reasons just outlined, the MoC is in the business of elucidating our concepts, then the nature of our concepts is a plausible place to look to better understand why the MoC leads to theoretical stalemates in the areas of philosophy where it plays a central role. In this section, I will argue that the MoC tacitly assumes a view of conceptual structure that psychology has decisively rejected. I will then, in the next section, argue that these mistaken assumptions about concepts explain why no theory has been, or ever will be, adequate by the standards that the MoC sets – and thus why MoC-based philosophy inevitably leads to theoretical stalemates.

²¹ I'll use SMALL CAPS to denote concepts/categories.

²² These points about concepts and categorization/classification broadly follow those made by Baz himself (2016a, pp. 111-112).

Specifically, the MoC assumes what is known in psychology as the **classical view of concepts** (a term coined by Smith & Medin 1981). According to the classical view, concepts (and their corresponding categories) are defined in terms of necessary and sufficient features. To take a stock philosophical example of a (putative) classical concept, the concept BACHELOR is defined as an “unmarried adult male”; the specified features – male, adult, and unmarried – are thought to be necessary and sufficient for being a bachelor.²³

Why think that the MoC assumes the classical view of concepts, even tacitly? One reason is simply that Western philosophy has long sought to come up with necessary and sufficient criteria for concepts of interest and has pursued this goal by testing the adequacy of proposed criteria against case-specific judgments; this type of project traces its roots at least as far back as Plato.²⁴ But another reason is that without the assumption of classical conceptual structure, it is difficult to make sense of the MoC’s joint commitment to (i) definition-like theories that attempt to provide necessary and sufficient criteria for their target phenomena and (ii) comparing theories of this sort against our intuitions (and thereby, as discussed above, against our concepts) as a means of assessing those theories’ adequacy and progressively revising them. That is, unless one assumes that our concepts have classical structure, it’s not clear why failing to capture a particular case-specific intuition should be a powerful mark against the adequacy of a theory consisting of

²³ Of course, even this supposed paradigm of a classically defined concept is subject to obvious counterexamples of unmarried adult males who are not (intuitively) bachelors: widowers, male clergy who are not permitted to marry, men who are in long-term romantic relationships but have not married their partners due to personal preferences or legal restrictions, etc.

²⁴ My focus in this paper is on the MoC (and its antecedents) as employed in Western philosophy, but Nagel (2013, p. 179) makes the interesting suggestion that similar uses of cases and intuitions also have a long history in Eastern philosophical traditions.

necessary and sufficient criteria, or why repeated testing and revision using our case-specific intuitions should be thought to get us closer to intuitively adequate criteria of this sort.

However, the problem with assuming the truth of the classical view of concepts is that there is very good evidence, in the form of an extensive literature from psychology, that the classical view is incorrect. This literature originated in the 1970s, when, for the first time, psychologists began empirically testing the classical view in earnest.²⁵ When they did so, the classical view failed spectacularly, with its key predictions systematically failing to be borne out by the experimental results. Contrary to what the classical view predicts, natural²⁶ concepts defined in terms of necessary and sufficient criteria are, in fact, vanishingly rare. Rather than there being a set of necessary and sufficient features shared by all and only members of a given concept/category (e.g., GAME, CHAIR), members are related by *family resemblance*, with overlapping sets of features shared between members but no set of

²⁵ Early theoretical doubts, as well as the suggestion that category members are related by family resemblance rather than necessary and sufficient features (subsequently borne out by empirical research – see below in main text), came from Wittgenstein (1953/2009).

²⁶ Psychologists (e.g., Rosch 1973, Mervis & Rosch 1981) studying concepts/categories distinguish between ‘natural’ concepts/categories (roughly, those that develop through various forms of experience with the world, such as perception and social learning) and ‘artificial’ concepts/categories (roughly, those that are imposed by stipulation, such as logical and mathematical concepts/categories defined as part of formal systems). The psychological work that I’m discussing in this section seeks to understand the structure of natural concepts; as such, when I, and psychologists such as those cited in this section, say that there is strong evidence that natural concepts do not have classical structure, this is compatible with the fact that artificial concepts may often be classical (e.g., TRIANGLE). Having said that, since my focus, as well as that of the psychological literature, is on natural concepts/categories – which comprise the vast majority of the concepts/categories that we use – I will generally just say ‘concepts’ instead of including the ‘natural’ qualifier each time.

defining features common to all (and only) those members (e.g., Rosch & Mervis 1975). And rather than concept/category membership being an all-or-nothing matter in which there is a sharp division between full-fledged members (items that possess the necessary and sufficient features) and full-fledged non-members (items that do not possess those features), concepts/categories have a graded structure in which there is a gradual transition from more typical to less typical items (e.g., Rosch 1975, Rosch & Mervis 1975).

This work has led to the all but complete rejection of the classical view in psychology. Writing in the early 1980s, Smith and Medin (1981, p. 33) remarked that “psychologists are currently forsaking the classical view in droves.” Two decades later, with the classical view’s abandonment by psychologists having played out, Murphy (2002, p. 16) described the view’s demise in the following unequivocal terms: “The groundbreaking work of Eleanor Rosch in the 1970s essentially killed the classical view, so that it is not now the theory of any actual researcher in this area.”

The tale of the classical view’s fall is one that can be found in many psychology textbooks, and in articles and books such as those cited in this section. Thus, for reasons of space – and, more generally, because as just described, the rejection of the classical view is about as close to a total consensus as one will find in psychology – I’m omitting detailed discussion of the empirical work that led to the classical view’s downfall (as well as the alternative models of conceptual structure that psychologists have developed in its place).²⁷ Instead, I will simply take the rejection of the classical view as a given and focus on

²⁷ Smith & Medin (1981, Ch. 3) is the seminal review of the empirical work that led to the fall of the classical view. See also Mervis & Rosch (1981) and Murphy (2002, Ch. 2). For overviews of psychological research on concepts/categories, see Murphy (2002), Machery (2009), and Goldstone et al. (2017).

examining the implications that this has for the MoC. This will be my project in the remainder of the paper.

4. From Psychology to Philosophy

The key implication of the classical view's rejection is simply this: the MoC is committed to two requirements for theories that are incompatible with one another and therefore cannot succeed by the very standards that it sets for itself. On the one hand, theories are expected to provide necessary and sufficient criteria for phenomena/concepts of interest (knowledge, morally right action, moral responsibility, etc.). On the other hand, theories are expected to match up with our case-specific intuitions as a core measure of adequacy. However, if the psychological consensus is correct and our concepts do not have classical structure (i.e., are not defined in terms of necessary and sufficient criteria), then no theory that has structure mandated by the former requirement will be able to meet the standard of intuitive adequacy mandated by the latter.

This mismatch provides a compelling explanation for the Problem of Counterexamples – the inevitability with which theories encounter cases where their verdicts fail to match up with our intuitions and are therefore deemed inadequate by the lights of the MoC. The cause of the Problem of Counterexamples is not that philosophers theorizing about a given phenomenon simply have not been clever enough to come up with a set of necessary and sufficient criteria that avoids counterexamples. Rather, the cause is structural: for any theory that proposes necessary and sufficient criteria for some phenomenon of interest, there will always be cases in which it gives intuitively incorrect verdicts, since the concepts underlying our case-specific intuitions about what is or is not an instance of the phenomenon do not have classical structure.

This can be seen most clearly by considering the notion of family resemblance briefly discussed above. Family resemblance, recall, is the idea – originally posited by Wittgenstein and later supported by the empirical research of Rosch and colleagues (e.g., Rosch 1975, Rosch & Mervis 1975, Mervis & Rosch 1981) – that concept/category members are related by overlapping sets of shared features rather than by possession of a common set of defining (necessary and sufficient) features. The claim that category members are related by family resemblance is not to be conflated with experimental philosophers’ claims regarding variation in case-specific judgments (see Section 1). The latter deals with variation in judgments about whether particular cases are or are not members of a given category (e.g., Person 1 judges that F is a member of Category C whereas Person 2 judges that F is not a member of C). By contrast, the family resemblance research deals with the internal structure of concepts/categories: what it shows is that even if there is no variation in case-specific categorization judgments (i.e., there is consensus about what is and is not a member of a given category), there is *still* no set of necessary and sufficient features that unify all of the things that we agree to be members of a given category (such as, to use Wittgenstein’s [1953/2009] example, the things that we agree are members of the category GAME).²⁸

If category members are related by family resemblance, then any philosophical theory proposing necessary and sufficient features for some phenomenon X – that is, features that are shared by all and only instances of X – will fail to fully match up with our case-specific intuitions, since there is no such set of features shared by all and only those

²⁸ The problems for necessary and sufficient criteria don’t end here, since as briefly mentioned in Section 3, such criteria will also fail to capture the graded nature of concept/category membership. See also Machery (2017, pp. 233-234).

things that we intuitively judge to be instances of X.²⁹ Crucially, this does not just apply to newly proposed theories: it applies to *any* theory consisting of necessary and sufficient criteria, no matter how many cycles of MoC-based testing and revision it undergoes. What this means, then, is that the Convergence Assumption is false: no amount of testing and revision via the MoC will give us a theory that meets the MoC's own standard of intuitive adequacy.

A further point about generality. Throughout the paper, I've used as examples phenomena/concepts such as knowledge, morally right action, causation, etc. that have been central objects of theorizing in their respective areas of philosophy for hundreds, and in some cases even thousands, of years. But the same points – regarding the impossibility of coming up with criteria that are simultaneously necessary and sufficient as well as intuitively adequate – apply just as much to less central/more recent philosophical phenomena/concepts. Indeed, the problem is a very general one afflicting *any* attempt to come up with intuitively satisfactory necessary and sufficient criteria for phenomena/concepts of interest. For instance, if a hypothetical “philosophy of furniture” employed the MoC in the service of devising theories of furniture phenomena/concepts such as chair and table, its practitioners would find themselves in exactly the kind of stalemates – and endless cycles of theory, counterexample, and revision – that we see for actual philosophical phenomena/concepts such as knowledge and moral rightness. It may just be that, because actual philosophical inquiry tends to be directed at difficult, complex targets, it becomes more difficult to notice – or at least easier to ignore – the fact that the

²⁹ To be clear, this doesn't mean that a theory consisting of necessary and sufficient criteria will give intuitively incorrect verdicts in *every* case with which it is confronted. Rather, what it means that there will always be some cases in which it fails to deliver intuitively correct verdicts (i.e., the theory will always be subject to counterexamples).

MoC would fail to reach a satisfactory conclusion about pretty much *any* phenomenon/concept at which it was directed.

What, then, is happening when we use the MoC to test and revise philosophical theories? According to the standard (optimistic) view of the MoC – a view that the Convergence Assumption tries to capture – we’re getting slowly but surely closer to intuitively adequate theories. The preceding discussion suggests a different (and more pessimistic) possibility: beyond a certain point – a point that has long since been passed in well-trod areas of philosophical inquiry – philosophical theories are not really getting better (at least by the standards of the MoC) through repeated cycles of testing/revision against our intuitions. Instead of moving from less adequate to more adequate theories (however slowly), we are simply moving from theories that are wrong in one way to theories that are wrong in a different way. In particular, we are moving from theories that are subject to one set of counterexamples to theories that are subject to another set of counterexamples. This lateral movement from one counterexample-prone theory to another is the sense in which MoC-based philosophy truly does fail to make progress even by its own lights.

The above analysis also helps to explain the entrenchment of theoretical disagreement in the many areas of philosophy where the MoC plays a central role. On the one hand, theories that do not provide necessary and sufficient criteria for phenomena/concepts of interest will be dismissed by many philosophers in these areas due to their perceived imprecision/vagueness and inability to give clear-cut case-specific verdicts. On the other hand, theories that do provide necessary and sufficient criteria – and thereby satisfy the demand for (a certain sort of) precision and clear-cut case-specific verdicts – will inevitably fail to be intuitively adequate and will be dismissed by many philosophers for that reason.

Faced with a landscape of theories that are all unsatisfactory in the ways just described, it is unsurprising that philosophers fail to arrive at any sort of consensus. In such

a state of affairs – where a field’s shared methods and assumptions (here, those that comprise the MoC) badly and perpetually underdetermine the answer to the question “which theory is best?” – one’s theoretical allegiances on a given topic, and the overall shape of theoretical allegiances among researchers in the field, will be determined less by standards that are shared and more by standards that are idiosyncratic: prior beliefs and commitments, relationships and personalities (mentors, colleagues, etc.), gut feelings, and so forth.³⁰

Such factors do, of course, also influence theoretical allegiances in other academic disciplines – including the sciences, where there are far fewer worries about persistent disagreement and lack of progress. The difference is that the shared methods and assumptions in most scientific disciplines – ranging from natural sciences like physics and biology to cognitive sciences like neuroscience and psychology – generally provide a stronger guiding hand than they do in philosophy. As a result, the relative influence of idiosyncratic factors is diminished, and consensus is more likely, especially over medium-to-long timescales.³¹

5. (Some) Defenses and Modifications of the MoC

In this section, I consider two possible responses to the pessimistic picture of the MoC painted in the preceding sections. The first appeals to the role of thought experiments and conceptual analysis in science as the basis for optimism about the MoC, while the

³⁰ For discussion of how a lack of shared standards can be an obstacle to philosophical progress, see Wilson (2017). For more on the influence of prior beliefs/commitments, see Chalmers (2015, p. 22).

³¹ For related points, albeit framed in terms of philosophical methods generally rather than the MoC specifically, see Daly (2017, pp. 36-38).

second contends that the MoC becomes viable with some modifications to its requirement of intuitive adequacy. I argue that neither of these responses succeeds.

5.1. Thought Experiments and Conceptual Analysis in Science

The contrast drawn at the end of the previous section suggests a sharp division between the methods of philosophy and those of the sciences. But of course, any such tidy division breaks down in various ways. Of particular relevance to this paper's topic are the following two. First, thought experiments may be used by scientists in an effort to gain insight into scientific matters.³² Second, philosophers of science may attempt to analyze concepts of scientific interest (which may themselves overlap with concepts of philosophical interest, as in the case of concepts such as LIFE and COGNITION).

There are various ways that these observations could be used to challenge my criticisms of the MoC. The general idea is as follows. The criticisms of the MoC developed in earlier sections are motivated, in significant part, by that method's persistent failure to resolve philosophical questions, as reflected in the neverending cycle of theory, counterexample, and revision into which MoC-based philosophical theorizing inevitably

³² The productive role of thought experiments in science is usually illustrated by pointing to the history of physics, which is filled with famous thought experiments – from figures such as Galileo, Newton, Maxwell, Einstein, and Schrödinger – that provided insight into existing theories, presented puzzles in need of solution, and/or inspired the development of new theories. And as Lakatos (1976) argues, an MoC-like process – involving mathematical conjectures being subjected to a search for counterexamples – has been an important driver of progress in the history of mathematics. It is, however, worth noting – although further examination of the issue lies well beyond the scope of this paper – that while examples of influential thought experiments in physics abound, their prevalence and contribution to other sciences is significantly less clear-cut (see, e.g., Snooks 2006, Swan 2009).

falls. However, if the MoC has a productive history in (at least some of) the sciences, then perhaps this pessimism about the MoC is unwarranted.

One response to these complications is simply to emphasize that my target is the *philosophical* MoC, as opposed to uses of thought experiments (cases, intuitions, etc.) that appear in other disciplines. (Baz [e.g., 2015, 2016a, 2016b], for instance, frequently adds the ‘philosophical’ qualifier in his critical discussions of the MoC.) The distinction is not merely semantic. Well-known uses of thought experiments in the sciences – particularly physics (see note 32) – bear little resemblance to the philosophical MoC that is my target in this paper. The productive uses of thought experiments in the history of physics occur against a backdrop of, and are meaningfully constrained by, a substantial body of shared formal, empirical, and theoretical commitments. Thought experiments and intuitions, while undoubtedly sometimes useful in physicists’ inquiries, are not the primary evidence upon which assessments of theories’ adequacy are based, nor are they being used to construct the sorts of definition-like theories that are characteristic of the MoC in philosophy. Physicists use thought experiments to help them devise better physical theories, not to devise and test the adequacy of necessary and sufficient criteria for concepts such as PHYSICAL.

What about efforts that *do* explicitly aim to analyze scientific concepts? Fully addressing the status of such projects would require far more space than I have available here, but it is a topic that is examined elsewhere against the backdrop of conceptual/definitional debates at the intersection of philosophy and the cognitive sciences (Kavanagh & Suhler 2015, Suhler & Kavanagh in preparation). The (very) short version of that account is as follows. Philosophers’ meticulous attention to definitions and distinctions can indeed be an asset in some scientific contexts, such as when attempts to compare and synthesize scientific findings are mired in confusion due to the same term/concept being used in different ways by different subdisciplines or even by different research groups within a subdiscipline. Such philosophical endeavors are at their best when they proceed by

looking at how the relevant term/concept is used in the actual theories and practices of the relevant sciences and are open to the possibility that the uses of the term in different contexts may not be fully reconcilable by a single tidy definition. By contrast, to the extent that efforts to analyze scientifically important concepts resemble the philosophical MoC in their dogmatic insistence on unifying all applications of the term/concept under a single (necessary and sufficient) definition and/or the heavy evidential weight that they accord to folk intuitions, we should be quite wary of them. Such projects are likely to devolve into the same sorts of stalemates as their more traditional philosophical counterparts and to have little practical or theoretical upside for the relevant science (or the philosophy thereof) – particularly when they are directed at the sorts of high-level concepts that are the most seductive targets for definition-seeking.³³

5.2. Intuitive Adequacy and Reflective Equilibrium

The central argument of this paper – laid out in Section 4 – is that the MoC is built around two commitments that are incompatible with one another: (C1) theories that give necessary and sufficient criteria for phenomena/concepts of interest and (C2) theories that match up with our case-specific intuitions as a core measure of adequacy. The incompatibility of these commitments is what makes it impossible for any theory to be deemed adequate by the lights of the MoC, which in turn helps to explain the entrenched theoretical disagreement and lack of progress in the many areas of philosophy that employ the MoC.

³³ Machery (2012) makes similar points in his critical discussion of attempts to define ‘life’, as does Allen (2017) for attempts to define ‘cognition’.

One possible response to this line of argument is that commitments C1 and C2 are only truly incompatible on an excessively strong conception of intuitive adequacy. And indeed, my earlier discussion at times may have given the impression that the MoC requires nothing less than that a theory of X match up with every single case-specific intuition that we have about X. However, case-specific intuitions can vary in their strength, centrality, and other merits; theories and their merits can likewise vary along numerous dimensions. Given this, it's plausible to think that what happens in instances of conflict between intuition and theory should depend on the relative strength of these merits in the particular instance at hand. On such an approach, then, a theory need not always be discarded when it conflicts with a case-specific intuition; on some occasions, such as when the intuition is weak or unclear, we may choose to keep the theory and discard the intuition.

What such a response would be invoking, then, is something like **reflective equilibrium** as a way of deciding whether intuition or theory should win out in situations where they conflict (i.e., when one is faced with a potential counterexample to a theory). This could be used to defend the MoC in the following way: even if, for the reasons laid out earlier in the paper, no theory consisting of necessary and sufficient criteria for X can match up with all of our case-specific intuitions about X, such a theory may be able to match up with the subset of intuitions that, after subjecting the matter to reflective equilibrium, we judge worth keeping (because they are particularly strong, or are elicited by cases we regard as particularly central, or whatever else). In this way, reflective equilibrium – and its more flexible picture of how to reconcile conflicts between general principles and specific judgments– may seem to offer a way of making commitments C1 and C2 compatible after all (viz., by making C2's standard of intuitive adequacy less stringent).

The reflective equilibrium approach just sketched may seem, at first glance, to be a very natural and promising avenue for salvaging the MoC. However, there are a number of reasons to doubt the prospects of this strategy.

One source of doubt is historical: optimism about reflective equilibrium as a way of salvaging the MoC is badly at odds with the actual track record of reflective equilibrium in philosophy. Despite its *prima facie* plausibility as a framework for reconciling competing considerations, the method of reflective equilibrium has done little or nothing to ameliorate the Problem of Counterexamples or pervasive theoretical disagreement since it gained widespread currency in the early 1970s.³⁴ The same disagreements that loomed large prior to reflective equilibrium's introduction have continued to dominate their respective areas of philosophy, and new topics (or sub-topics) of philosophical disagreement have risen up to join the existing ones.

Why has reflective equilibrium done so little to resolve theoretical disputes in philosophy, and more specifically to help philosophers satisfactorily adjudicate conflicts between theories and intuitions? A key reason is that what constitutes reflective equilibrium may be quite different from one person to the next – what Stich (1988) terms the *problem of cognitive diversity*. Adapted to the context of the present paper, the crux of the problem of cognitive diversity is that different individuals may have widely divergent ways of achieving coherence between theory and intuition. To use an extreme example, reflective equilibrium for a card-carrying consequentialist may be a state in which all intuitive counterexamples to consequentialism are discarded, while reflective equilibrium for a card-carrying deontologist may be a state in which all intuitive counterexamples to deontology are discarded. To be sure, most philosophers' equilibrium points on a given matter are not likely to be quite so extreme as this toy example. But this still leaves ample scope for diversity in equilibria between individuals, especially given that the theories and intuitions being considered will often involve (e.g.) difficult/borderline cases and topics where

³⁴ The label 'reflective equilibrium', along with an influential statement of the method itself, comes from Rawls (1971). For precursors, see Rawls (1951) and Goodman (1955).

individuals have prior commitments/allegiances. The upshot of this is that reflective equilibrium is unlikely to provide any general, univocal perspective from which to determine which intuitions our theories really do or do not need to capture, since there are numerous, mutually incompatible ways of achieving equilibrium between theories and intuitions – and no non-question-begging way of saying which is correct (a point also emphasized, in a different dialectical context, in Stich’s [1988] original article).

Furthermore, it’s not clear that the idea that we should discard or discount questionable intuitions – such as those that are weak/unclear or based on dubious/contrived/peripheral cases – even matches up all that well with philosophical practice.³⁵ Philosophical thought experiments routinely involve all sorts of contrivances – extreme coincidences, science fiction technologies, bizarre environments, implausible stipulations, and so forth – yet this does not prevent such cases (or, more precisely, the intuitions they elicit) from being regarded by many philosophers as powerful evidence for or against theories’ adequacy. The use of such cases/intuitions has not entirely escaped criticism, of course: one particularly well-known line of criticism is Dennett’s (e.g., 1984, 1991) contention that “intuition pumps”, while potentially valuable if used judiciously, are often abused by philosophers.³⁶ But such worries have done little to stop the use of contrived/unrealistic/etc. thought experimental cases in philosophy, owing (to echo the previous paragraph) to the fact that there’s no clear, agreed-upon standard for determining

³⁵ For instance, writing about cases of theory-intuition conflict in epistemology, Weatherson (2003, p. 1) remarks that there have been “literally dozens of cases where an interesting account of knowledge was jettisoned because it clashed with intuition about a particular case.” For more general discussion of philosophers’ tendency to assume that theories must yield to intuitions, see Mizrahi (2015).

³⁶ More recent criticisms of the MoC, such those from Machery (2017, Ch. 3) and Mizrahi (2014), have taken a similar line, arguing that the unusual nature of many philosophical cases renders the intuitions they elicit unreliable.

which cases are mere intuition pumps (in the pejorative sense) and which are genuinely probative.

The upshot of this is that appealing to reflective equilibrium or otherwise trying to place a principled limit on the intuitions that our theories must capture is unlikely to resolve the Problem of Counterexamples or the widespread theoretical disagreement to which it leads. What about instead weakening C1 – the requirement that theories must provide necessary and sufficient criteria for their target phenomena/concepts? While I think that theories not built around necessary and sufficient criteria are well worth considering in many areas of philosophy (see note 16), this approach is not a viable one for a defender of the MoC.

The reason is that as discussed in Section 2, necessary and sufficient criteria are what allow us to precisely identify a theory's verdict in any sufficiently well-described case. If practitioners of the MoC no longer require that theories provide necessary and sufficient criteria for their target phenomena/concepts, or if a theory's "core" necessary and sufficient criteria are allowed to be accompanied by various exceptions and alternative criteria, they create problems of uncertainty and indeterminacy about what the theory says in particular cases (especially when those cases are difficult ones with many competing considerations). Thus, besides being unsatisfying for the sorts of theorists who are drawn to the MoC – and its promise of theories that provide clear, intuitively satisfying case-specific verdicts – in the first place, weakening or abandoning the requirement that theories consist of necessary and sufficient criteria would undermine the core methodology of the MoC: without theories that yield precise case-specific verdicts that can then be compared against our intuitions, the MoC cannot get off the ground (see Section 2).

6. Philosophy without the MoC?

If the MoC has the deep flaws that I'm suggesting it does, then shouldn't we seek out alternative methods to take its place? For reasons I'll elaborate on momentarily, I think that the answer to this question is 'yes'. But as Daly (2017) observes, we should be cautious about pinning all our hopes for the resolution of philosophical disagreement on radical methodological change:

Does the answer lie in the devising of some 'brave new method' that will replace or supplement our existing methods? The possibility of methodological innovation should not be ruled out, but even so, we should not hold out for the prospect of it making a breakthrough in our philosophical fortunes... Issues about the new method's justification, its reliability, and its standing relative to our other methods would need to be debated and resolved, and, on the face of it, there is no reason to think that these issues would be any more tractable in the case of any new method than they have been in the case of any of the methods that we have already got. (Daly 2017, p. 36)

I'm sympathetic to the worries Daly raises³⁷; indeed, the previous section's discussion of reflective equilibrium is an example of how supplementing one method (the MoC) with another (reflective equilibrium) doesn't really do anything to resolve disagreement and instead at most pushes it back a level.³⁸ However, the fact that any philosophical method

³⁷ For similar worries, see Chalmers (2015, §5).

³⁸ One possibility, suggested by an anonymous referee, is that the failure of reflective equilibrium to meaningfully clarify the MoC is due to reflective equilibrium itself relying on intuitions – for instance, intuitions about the relative strength of the different considerations one is trying to bring into

(and its assumptions, etc.) is going to be contested at some level does not mean that some methods (and their assumptions, etc.) cannot be better or worse than others. And there are reasons to think that the MoC is an especially problematic method around which to organize philosophical inquiry.

For starters, the key empirical assumption on which the MoC depends – namely, that our concepts have classical structure – is one that our best science gives us strong reason to think is false (see Section 3). Of course, what role “our best science” should play in our philosophizing is itself one of the many topics about which philosophers have long disagreed. But if we want our philosophizing to be even minimally constrained by facts about how the world is, it seems quite bizarre to employ a method that rests on a theory of concepts that we have very good scientific reason to think is incorrect; because of its dubious foundations, such a method is more likely to mislead our theorizing than to guide us toward good theories.³⁹

equilibrium – and therefore not really constituting a distinct level of analysis or source of evidence. I think that this is an interesting and plausible suggestion, but it’s not one that I will not pursue further here since I want to keep the notion of intuition at work in this paper as simple as possible (see note 9).

³⁹ Machery (2017) offers support for this worry, linking the failings of traditional philosophical attempts at both prescriptive (pp. 213-214) and descriptive (pp. 233-234) conceptual analysis to philosophers’ assumption that concepts must be characterized in terms of necessary and sufficient features. He does this in the service of setting up his own positive proposals for the goals and methods of philosophy, most notably a thoroughgoingly empirical approach to descriptive conceptual analysis that he terms the “method of cases 2.0” (pp. 239-240). Nothing in my paper’s argument threatens Machery’s “method of cases 2.0”, since his project explicitly rejects the assumption of classical conceptual structure on which traditional philosophical engagement with

In line with this, there are real costs that come with continuing to organize many areas of philosophical inquiry around the MoC. Perhaps the most important cost comes in the form of the rejection or marginalization of theories on spurious grounds. The very thing that makes the MoC so attractive – the idea that we can get theories that are both extremely precise in their case-specific verdicts (through necessary and sufficient criteria) and extremely intuitively adequate (through repeated testing and revision of such criteria against our case-specific intuitions) – is also what makes it so problematic, since it leads to theories being rejected because they fail to meet this impossible standard.⁴⁰ In this way, MoC-based philosophy not only often fails to make progress by its own lights (see Section 4)

cases and intuitions – i.e., what I’ve been calling the MoC and what Machery calls the “method of cases 1.0” – has been based.

With that said, there are important differences between my argument and Machery’s (2017) that are worth highlighting. Machery’s argument against the MoC, as developed in the first six chapters of his book, is an (especially thorough and compelling) experimental philosophy-style argument from variation in case-specific intuitions, with the points about classical conceptual assumptions only brought in briefly in the final chapter as foils for his positive projects. By contrast, my argument shows how the non-classical structure of our concepts is enough, on its own, to doom the MoC, without any need for further claims about the existence of variation in case-specific intuitions. Thus, while I have no significant objections to Machery’s variation-based argument against the MoC – and, given his critical remarks about the flawed assumptions of traditional conceptual analysis, I doubt he would have significant issues with mine – they are best seen as complementary attacks on the MoC.

⁴⁰ There may, of course, turn out to be good philosophical reasons to reject some or even all of the theories that fall short of the MoC’s standards, but the fact that they don’t meet those (impossible) standards is not such a reason.

but may even work against genuine theoretical progress by leading to the rejection or marginalization of otherwise promising theories on spurious grounds.⁴¹

Another cost is the substantial philosophical effort expended under the assumption that counterexamples tell us something profoundly important about a theory's (in)adequacy: devising cases, debating their status (e.g., attempting to explain away intuitions that are problematic for one's theory), revising theories to avoid counterexamples, devising new counterexamples to the revised theory, and so forth. However, once we realize that the mere existence of a counterexample doesn't necessarily tell us anything all that interesting about an otherwise plausible theory's merits relative to its rivals (since *every* theory that gives necessary and sufficient criteria will have cases where it fails to match our intuitions) or provide a stepping stone toward a version of the theory that is not subject to counterexamples (since no matter how much revision the theory undergoes in response to counterexamples, there will still be cases where it fails to match our intuitions), the value of this counterexample-focused approach diminishes.⁴²

The upshot of this is that while cases and intuitions may still have some useful role to play in philosophical theorizing, they cannot play the central evidential role that the MoC, and the many areas of philosophy that employ it, have long taken them to. And to persist under the assumption that they can play this role is to continue to use a method that we have good reasons – in the form of the MoC's track record and the considerations set out in this paper, as well as arguments from experimental philosophers (e.g., Machery 2017, Stich

⁴¹ Recall, for instance, Weatherson's (2003, p. 1) remark that dozens of promising theories of knowledge have been rejected simply because they failed to match up with our intuitions in a particular case (see note 35).

⁴² For further critical discussion of the weight accorded to counterexamples, see Mizrahi (2015) and Weatherson (2003).

2013) and other critics of the MoC (e.g., Baz 2015, 2016a, 2016b, 2017, Mizrahi 2014, 2015) – to think is incapable of delivering what it promises.

What, then, is the way forward? The attraction of the MoC is that we can have our theoretical cake and eat it too, in the form of theories that both provide necessary and sufficient criteria for phenomena of interest (knowledge, morally right action, free will, cognition, etc.) and do an excellent job of capturing our case-specific intuitions about those phenomena. Once we see that this is likely nothing more than a pipe dream, we will have to make hard choices about what form we want our theories to take and what role we want intuitive adequacy to play in evaluating those theories.⁴³ This will not by itself resolve theoretical disagreements in philosophy, of course, but it will at least give us a clearer picture of what we can expect from our theories/methods and what we are really disagreeing about. And that is itself a form of progress.

References

Allen, C. (2017). On (not) defining cognition. *Synthese*, 194(11), 4233–4249.

Andow, J. (2015). How Distinctive Is Philosophers' Intuition Talk? *Metaphilosophy*, 46(4–5), 515–538.

⁴³ For instance, we might, among many other possibilities, reject the expectation of necessary and sufficient definition for some central philosophical concepts (see note 16) or uses cases/intuitions as the basis for psychologically rigorous descriptive study of the structure of our philosophical concepts (see note 39).

Ashton, Z., & Mizrahi, M. (2018). Intuition Talk is Not Methodologically Cheap: Empirically Testing the “Received Wisdom” About Armchair Philosophy. *Erkenntnis*, 83(3), 595–612.

Baz, A. (2015). Questioning the Method of Cases Fundamentally – Reply to Deutsch. *Inquiry*, 58(7–8), 895–907.

Baz, A. (2016a). Recent Attempts to Defend the Philosophical Method of Cases and the Linguistic (Re)turn. *Philosophy and Phenomenological Research*, 92(1), 105–130.

Baz, A. (2016b). On going (and getting) nowhere with our words: New skepticism about the philosophical method of cases. *Philosophical Psychology*, 29(1), 64–83.

Baz, A. (2017). *The Crisis of Method in Contemporary Analytic Philosophy*. Oxford: Oxford University Press.

Bealer, G. (2000). A Theory of the A Priori. *Pacific Philosophical Quarterly*, 81(1), 1–30.

Blackford, R., & Broderick, D. (Eds.). (2017). *Philosophy's Future: The Problem Of Philosophical Progress*. Hoboken, NJ: Wiley-Blackwell.

Brennan, J. (2010). Scepticism about Philosophy. *Ratio*, 23(1), 1–16.

Cappelen, H. (2012). *Philosophy Without Intuitions*. Oxford: Oxford University Press.

Chalmers, D. J. (2014). Intuitions in philosophy: a minimal defense. *Philosophical Studies*, 171(3), 535–544.

Chalmers, D. J. (2015). Why Isn't There More Progress in Philosophy? *Philosophy*, 90(1), 3–31.

Christensen, D. (2007). Epistemology of Disagreement: The Good News. *Philosophical Review*, 116(2), 187–217.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19.

Climenhaga, N. (2018). Intuitions are Used as Evidence in Philosophy. *Mind*, 127(505), 69–104.

Daly, C. (2017). Persistent Philosophical Disagreement. *Proceedings of the Aristotelian Society*, 117(1), 23–40.

Dennett, D. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little Brown.

Deutsch, M. (2010). Intuitions, Counter-Examples, and Experimental Philosophy. *Review of Philosophy and Psychology*, 1(3), 447–460.

Deutsch, M. (2015). *The Myth of the Intuitive*. Cambridge, MA: MIT Press.

Frances, B. (2017). Extensive Philosophical Agreement and Progress. *Metaphilosophy*, 48(1-2), 47-57.

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23), 829-839.

Fumerton, R. (2010). You Can't Trust a Philosopher. In R. Feldman & T. A. Warfield (Eds.), *Disagreement* (pp. 91-111). Oxford: Oxford University Press.

Gettier, E. L. (1963). Is Justified True Belief Knowledge? *Analysis*, 23(6), 121-123.

Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2017). Categorization and Concepts. In J. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive neuroscience, Fourth Edition, Volume Three: Language & Thought* (pp. 275-317). New York: Wiley.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy*, 83(5), 291-295.

Jackson, F. (2009). Thought Experiments and Possibilities. *Analysis*, 69(1), 100-109.

Kavanagh, L., & Suhler, C. (2015). What is the role of conceptual analysis in cognitive science? In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.). *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 1057-1062.

Kornblith, H. (2013). Is Philosophical Knowledge Possible? In D. E. Machuca (Ed.), *Disagreement and Skepticism* (pp. 260–276). New York: Routledge.

Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kuntz, J. R., & Kuntz, J. R. C. (2011). Surveying Philosophers About Philosophical Intuition. *Review of Philosophy and Psychology*, 2(4), 643–665.

Lakatos, I. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.

Machery, E. (2009). *Doing Without Concepts*. New York: Oxford University Press.

Machery, E. (2012). Why I stopped worrying about the definition of life... and why you should as well. *Synthese*, 185(1), 145–164.

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(3), B1–B12.

Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, 32(1), 89–115.

Mizrahi, M. (2014). Does the Method of Cases Rest on a Mistake? *Review of Philosophy and Psychology*, 5(2), 183–197.

Mizrahi, M. (2015). Don't Believe the Hype: Why Should Philosophical Theories Yield to Intuitions? *Teorema: Revista Internacional de Filosofía*, 34(3), 141–158.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Nado, J. (2016). The intuition deniers. *Philosophical Studies*, 173(3), 781–800.

Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research*, 85(3), 495–527.

Nagel, J. (2013). Defending the Evidential Value of Epistemic Intuitions: A Reply to Stich. *Philosophy and Phenomenological Research*, 87(1), 179–199.

Parfit, D. (2011). *On What Matters* (Vol. 1). New York: Oxford University Press.

Pereboom, D. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Pust, J. (2001). Against Explanationist Skepticism regarding Philosophical Intuitions. *Philosophical Studies*, 106(3), 227–258.

Rawls, J. (1951). Outline of a Decision Procedure for Ethics. *The Philosophical Review*, 60(2), 177–97.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, 4(3), 328–350.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7(4), 532–547.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605.

Searle, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3, 417–457.

Shand, J. (2017). Philosophy Makes No Progress, So What Is the Point of It? *Metaphilosophy*, 48(3), 284–295.

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Snooks, R. J. (2006). Another scientific practice separating chemistry from physics: thought experiments. *Foundations of Chemistry*, 8(3), 255–270.

Stich, S. (1988). Reflective equilibrium, analytic epistemology and the problem of cognitive diversity. *Synthese*, 74(3), 391–413.

Stich, S. (2013). Do Different Groups Have Different Epistemic Intuitions? A Reply to Jennifer Nagel. *Philosophy and Phenomenological Research*, 87(1), 151–178.

Stoljar, D. (2017). *Philosophical Progress: In Defence of a Reasonable Optimism*. Oxford: Oxford University Press.

Suhler, C., & Kavanagh, L. (in preparation). Does Cognitive Science Need a ‘Mark of the Cognitive’?.

Swain, S., Joshua, A., & Weinberg, J. M. (2008). The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research*, 76(1), 138–155.

Swan, L. S. (2009). Synthesizing insight: artificial life as thought experimentation in biology. *Biology & Philosophy*, 24(5), 687–701.

Sytsma, J. (2010). The Proper Province of Philosophy: Conceptual Analysis and Empirical Investigation. *Review of Philosophy and Psychology*, 1, 427–445.

Turri, J., & Friedman, O. (2014). Winners and losers in the folk epistemology of lotteries. In J. Beebe (Ed.), *Advances in experimental epistemology* (pp. 45–69). London: Bloomsbury.

Weinberg, J. M. (2014). Cappelen between rock and a hard place. *Philosophical Studies*, 171(3), 545–553.

Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and Epistemic Intuitions. *Philosophical Topics*, 29(1/2), 429–460.

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, T. (2007). *The Philosophy of Philosophy*. New York: Routledge.

Wilson, J. (2017). Three Barriers to Progress in Philosophy. In R. Blackford & D. Broderick (Eds.), *Philosophy's Future: The Problem of Philosophical Progress* (pp. 91–104). Hoboken, NJ: Wiley-Blackwell.

Wittgenstein, L. (1953/2009). *Philosophical investigations* (Rev. 4th ed.). Trans. G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte. Malden, MA: Wiley-Blackwell.