Chapter 5

Thought Experiments and the (Ir-) relevance of Intuitions in Philosophy

Daniel Cohnitz

1 THE TWO DILEMMAS OF METAPHILOSOPHY

There is a growing interest in metaphilosophy, the study of the methodology of philosophy. A central topic in recent contributions to metaphilosophy is the role of intuitions in philosophy, especially their role as evidence for (or rather against) philosophical theories. Intuitions seem to play this role in conjunction with so-called "hypothetical cases" or "thought experiments". On the standard view, philosophy proceeds by considering hypothetical cases and forming intuitive judgements about these cases, which are then in turn used as evidence in favour or against philosophical theories. Metaphilosophy is then concerned with a general assessment of this practice: Are intuitive judgements indeed reliable evidence for philosophical theories? Experimental philosophers (e.g. Machery, 2017) argue that intuitions can empirically be shown to vary a lot and thus to be unreliable, while other metaphilosophers try to defend their use.

I have my doubts about the project of a general metaphilosophy, at least if it is understood as a kind of normative philosophy of science applied to philosophy. I believe that there are two dilemmas metaphilosophy on that conception is confronted with, which make it very unlikely that it can arrive at interesting and binding results. The reason is that philosophy (in a way to be explained below) is (and should be) too heterogeneous to allow for such results. I will explain the two dilemmas in this section. I also have my doubts about the particular focus on intuitions and the discussion of their evidential weight.

However, the current discussion of the role of intuitions seems to be evidence that there is a way around both dilemmas, and that the focus on intuitions is justified. Apparently, philosophy is not as heterogeneous as the prima

16028-0337dr2.indd 95 27-02-2020 19:49:41

facie considerations that I will offer below suggest. It seems that "hypothetical cases" play a role everywhere in philosophy, and intuitions play a role in hypothetical cases. Therefore, if we understand the role that intuitions play there, we actually can arrive at interesting and binding results for almost all areas of philosophy. The current metaphilosophical discussion often refers to this as "the method of cases". In section 2, I will explain what this method is supposed to be.

In the rest of my paper, I will argue that appearances are misleading here. Even though hypothetical cases, in a broad sense of the notion, play some role in many areas of philosophy, this role is not always the same. And for most of the roles we will identify in section 3, only one of them could give rise to the methodological worries about the evidential weight of intuitions (such as those raised by experimental philosophers that I alluded to above), because only one of these roles has room for an evidential use of intuitions.

Thus, even if hypothetical cases are often considered in philosophy (for some purpose or other), the fact that we might be able to say something interesting and binding about one role that hypothetical cases play does not mean that we have thereby said something interesting and binding about all areas of philosophy, or the dominant methodology, etc.

Now, the role that does have room for an evidential use of intuitions is the role of presenting hypothetical cases as counterexamples. Still, one might argue, this methodological function of hypothetical cases is common enough. So why should it not be possible to say something interesting and binding about this role and the evidential use of intuitions in this case?

As I will argue in section 4, there are at least three different models for how we might think about the way in which intuitive judgements provide "counterexamples" to a proposed theory. These different models are associated with different philosophical projects and motivated by certain assumptions about the nature of the subject matter of philosophical inquiry. On only one of these models would intuitions play an evidential role of the kind that seems to be assumed in the current metaphilosophical debate.

In section 5, I will discuss whether that latter model is often instantiated when hypothetical cases are presented as counterexamples, and I will offer an alternative conception of how we manage by way of presenting hypothetical cases to refute a theory, which does not refer to intuitions at all. Moreover, this alternative also has certain explanatory virtues over the idea that intuitions play a relevant role in the evaluation of hypothetical cases.

Section 6 will discuss the weight of the evidence I presented. I will argue that the evidence establishes the following claims: The abundance of hypothetical cases in philosophy is not evidence of a common and widespread methodology that crucially involves intuitions as evidence: (1) Most methodological roles of hypothetical cases have no room for any

role of intuitions. (2) Using hypothetical cases as counterexamples leaves room for intuitions, but we can distinguish at least three roles intuitions could play there, only one of which is an evidential role. (3) For the cases where this last model would apply, there is a plausible alternative story of how hypothetical cases work, that does not involve intuitions at all. Hence, my scepticism about metaphilosophy in general and its discussion of the special role of intuitions in philosophy in particular seem vindicated.

1.1 Interesting?

So, what are the two dilemmas I see for metaphilosophy? The first one has to do with the fact that philosophy is simply too heterogeneous to allow for a very interesting general methodology. Methodological questions arise relative to the aims or goals one has set oneself. They are questions of the type "What is the best way to achieve *X*?" But it is not clear at all that all areas of philosophy plausibly involve the same *X* here at any interesting level of abstraction. Let us see why. Perhaps the best candidate for philosophy's general goal is *to find the truth*. After all, philosophy is a compound of *φιλειν* and *σοφια*, and translates as *Love of Wisdom*, so should we not conceive of philosophy's aim in general to be cognitive?

First of all, it is not clear that all philosophy aims at truth. Is practical philosophy, ethics in particular, after discovering moral truth? If you are a non-cognitivist about ethics, you will dispute such assumption. And ethics is not the only area in which a non-cognitivist attitude might make sense. In those cases, the aim of philosophical inquiry might rather be seen in the development of normative proposals (for example, proposals to use certain concepts in certain refined ways, for all areas of philosophy who conceive of their enterprise as one of providing *explications* in Rudolf Carnap's sense of the word, cf. Carnap, 1967).

Second, even if we would agree that at least large parts of philosophy aim at truth, it is not clear that "truth" alone characterizes an aim that could determine a specific methodology. Different kinds of truths might require different methods for their discovery. Is our epistemic access to metaphysical truths the same as our access to conceptual or logical truths? Those who hold on to an analytic/synthetic distinction might doubt that it is. Logical truth is perhaps plausibly accessible *a priori*, but at least some metaphysical truths might be only *a posteriori* knowable.

Rejecting the *a priori*/*a posteriori* distinction improves the situation only slightly. Sociology, Linguistics, History, Particle Physics and Astronomy are all *a posteriori* disciplines, but they have very different methodologies, that display similarities only on a very high level of abstraction. The reason for

that is that their subject matter and our epistemic access to that subject matter is very different, even if it is all *a posteriori*.

It seems that in order to argue for a general methodology for philosophy, one would first need to answer a number of substantial philosophical questions in specific ways. Are moral questions cognitive? Are there metaphysical truths that are only *a posteriori* knowable? Is logic *a priori*? Are there any analytic truths, knowable on the basis of linguistic competence alone? When dealing with questions like these, one is already engaging with central topics in first-order philosophy, one is already *doing* philosophy. But how can one hope to formulate a methodology of philosophy, if that presupposes answers to questions like these?

Thus, the first dilemma is that the wider one wants to cast one's net in order to formulate a methodology for as many areas of philosophy as possible, the less interesting the methodological recommendations are going to be. Either we go for a discipline-wide methodology which is going to be trivial, or we chose a more substantial methodology, which then holds only for a very narrow subfield.

1.2 Binding?

The second dilemma is closely related. Let us assume for the sake of argument that there is a subsection of philosophy that on the one hand contains a good deal of philosophy (say, for the sake of an example, all of theoretical philosophy, excluding only applied ethics), and for which one was indeed able to identify a common subject matter (say, one endorses a philosophical view on which all of theoretical philosophy is interested in metaphysical truth). And let us further assume that for that subject matter, one finds and argues for a suitable methodology (perhaps one, according to which our epistemic access to metaphysical necessity and possibility is based in our general capacity to evaluate counterfactuals).

In this case, one might have found a methodology which is in equilibrium with a considerable amount of first-order philosophical views. Views about the subject matter of various sub-branches of theoretical philosophy, views about how metaphysical truth differs from, for example, conceptual truth or mere contingent empirical truth, views about the reducibility of metaphysical necessity and possibility to certain counterfactuals, views about the epistemological consequences such a reduction would have, and finally views about the unproblematic nature of our knowledge of counterfactuals.

The problem is that the view as a whole might be a considerably stable package, but this methodology will be unfounded for those who do not endorse all elements of that package, or who want to critically engage with some of those elements. At least, it will be in danger of being unfounded, since thus far, all

one has done is establishing that there is this *one* equilibrium between certain first-order philosophical views and a certain methodology.

Perhaps there is some interesting philosophy one can do in this tight corset of views. But the problem is that the normative proposal will only be binding for philosophers who accept to be limited by that corset of views. Thus, instead of having defended a methodology for philosophy, one will have defended a methodology for philosophers who share first-order commitments $C_1, \ldots, C_n$. Here the problem is practical: Does it make sense to formulate a methodology in philosophy, when it is binding only for a few confederates? Again we are confronted with a similar dilemma; we may go for a methodology that is binding for a larger group, but only at the cost of arriving at a methodology that is again trivial.

Another question is whether having a substantial and general methodology makes a lot of sense in philosophy. It is one thing to strive to choose methodological means that are not in conflict with one's theoretical commitments, for the sake of avoiding pragmatic contradictions. But it is a different project to formulate a normative philosophical methodology. If the latter requires a commitment to a certain set of first-order views, then it seems diametrically opposed to the spirit of philosophy as a maximally open inquiry in which not only everything is up for debate, but where actually everything is in fact debated. If that is what metaphilosophy is all about, then it is not clear what good metaphilosophy can do for philosophy.

2 THE METHOD OF CASES

However, this pessimism about metaphilosophy seems refuted by the discovery that there – after all – is a method that is common to all of philosophy. Joel Pust describes the method as follows (Pust, 2016):

> ***The Method of Cases***
>
> A philosophical theory is taken to be prima facie undermined by contradicting an intuition regarding a particular hypothetical case.

How widespread the method is in philosophy, and that it can be found in all its subareas, is usually confirmed by providing a list of prominent hypothetical cases from all these areas. Pust does that too (Pust, 2016):

> Such examples [Gettier case, Thomson's transplant case, Block's Chinese nation, Bromberger's flagpole case] could easily be multiplied to include teletransportation and fission cases in the literature on personal identity, preemption and epiphenomena cases in the literature on causation and explanation,

clairvoyance and evil demon cases in epistemology, Newcomb cases in decision theory, Frankfurt cases in the literature on free will, Twin-Earth and Swampman cases in the literature on mental content, Jackson's Mary case in the metaphysics of mind, trolley cases in applied ethics, experience machine cases in normative ethics, and many many others.

I want to argue in this paper that, although there are some methods in philosophy that fit under the description provided by Pust, these are not identical methods, and they are less prominent in philosophy than the list of hypothetical cases typically provided would suggest (in fact, those methods that fit comfortably under the description provided are somewhat idiosyncratic). There is an obvious problem with the kind of argument I would like to make. First of all, it seems that if methods have a common description, then unless they all fall under the description because the description is ambiguous, it should be fair to count them as being the same method according to that level of description. Second, in light of the first point, I seem to be also committed to the claim that there is one level of description that is the *proper* level for characterizing a method, such that descriptions at a higher level of generality would be too general to determine a method. Otherwise, one could respond to my criticism of the provided characterization of the method of cases by simply providing a more general characterization, which would then serve as a new candidate for a definition of the common method of philosophy.

I think one has to acknowledge (as I already did in the introduction) that there might well be a "method" that is common to all philosophy if this method is described in extremely general terms. However, I also believe that acknowledging the existence of this common method will not tell us anything interesting about philosophy, nor will the analysis of this method lead to interesting normative conclusions about how philosophy should be done.

However, both of the latter are supposed to be true of the method of cases. Rationalists, deflationists and experimental philosophers alike assume that the method of cases is described at a level of generality that allows to investigate what particular role intuitions play in the refutation of theories, and whether it can be justified that intuitions play that role. So I should clarify the aim of my paper accordingly: I will argue that there is no method of cases in philosophy that is in any interesting way widespread and would allow discussion of whether the use of intuitions is justified in general in that method.

In the next section, I will first show that the (indeed widespread) use of hypothetical scenarios in philosophy does not correspond to an equally widespread method. Hypothetical cases serve several very different functions in philosophy.

3 HYPOTHETICAL CASES IN PHILOSOPHY

Critics as well as advocates of the use of intuitions in philosophy usually point to long lists of hypothetical cases that we find in philosophy and suggest that this establishes that there is a common method in philosophy: the method of cases. However, the first thing to note is that hypothetical cases serve several different functions in philosophy. Here is a quick (and probably incomplete)[14] overview.

3.1 Analogies

Often, hypothetical cases are used as analogies and thus as a means of illustration. This is one way to interpret Plato's cave story. Plato is describing a hypothetical situation of people chained in a cave and looking at shadows that they confuse for the real things, in order to illustrate how, on his account, our own epistemic predicament keeps us from appreciating what is real. The point of the story is to illustrate our epistemic predicament by analogy with the predicament of those in the cave. On this interpretation, Plato is not testing a theory. He is not even quite presenting a thought experiment, because – like ordinary experiments – thought experiments consider scenarios from the domain of things at issue. But what is at issue here is Plato's theory of our ordinary knowledge. The situation in the cave is not in the domain of that theory; the point of the story is that the situation in the cave is just in certain ways analogous to cases of ordinary knowledge.

Let us make that explicit by defining a thought experiment as the consideration of a hypothetical case from the domain of things at issue. This definition should make the notion of thought experiment as flexible as our notion of an

14 One could add a further category that one might call "functional thought experiments". This category of thought experiments is again a bit of an outlier, just as analogies are. However, the cases that fall into this category are often called "thought experiments" and it is worth having a quick look, because they actually do cause certain epistemological concern. Philosophers of science identified these "functional" thought experiments (Stöltzner, 2004). Such thought experiments do not originate from outside a theory in order to test or illustrate the theory, but they are themselves part of the theory. The theory itself prescribes the contemplation of a certain thought experiment in its application conditions.

It is easiest to describe what is meant by this with a philosophical example. Some theories in practical philosophy prescribe that in order to determine whether a certain action is morally permissible one should first perform certain hypothetical considerations. For example, such a theory might suggest to hypothetically consider the evaluation of the consequences of the action from the point of view of every person that will be affected by the consequences of that action. This is a hypothetical exercise, and it is about the domain of things at issue. It is also epistemically dubious. First of all, such thought experiments must lead to reliable outcomes in order to be useful for their purpose. That means, they must actually lead to good estimates of how the consequences of a certain action would be evaluated by those affected, since otherwise the whole exercise would be pointless. However, for all we know about human psychology, these thought experiments are highly unlikely to lead to reliable predictions.

experiment is. Experiments do not always test – sometimes they *illustrate* (just think about the experiments your chemistry teacher performed in class), or they *explore* the behaviour of a system that you like to learn more about (perhaps yet in absence of any theory).

3.2 Illustrations

Like many mere analogies also thought experiments often only illustrate a point. Which thought experiments only serve such illustrative purposes might be controversial (for reasons to be discussed below), but here is a nice example from physics: imagine what would happen to the earth if the sun were suddenly to explode. According to the Newtonian theory, the earth would immediately depart from its usual elliptical orbit, while Einstein's theory predicts that the earth would stay in its usual orbit for another eight minutes, the time it takes for light to travel from the sun to earth. Here, the hypothetical cases (the sun suddenly explodes) are used to illustrate the difference between Newton's and Einstein's theory. The two theories make significantly different predictions about a certain extreme case.

Of course, such illustrations also work in the case of just a single theory. Physics textbooks are full of examples that illustrate how a theory would apply, with the intention to explain the content of the theory to the reader. What matters for the success of such thought experiments is whether they get the intended message across, whether the reader indeed achieves a better grasp of the theory so illustrated. What does not matter much for this is whether the outcome is significantly "intuitive" in any sense. It does not matter either whether the scenario described is in any sense really possible (e.g. whether it is physically possible that the sun simply disappears without the event causing this (an explosion, etc.) having any relevant effects of its own).

3.3 Puzzle Cases

These thought experiments typically have the features that Cappelen identified in his (2012): They function as a "fact focuser" in that they "draw our attention to a philosophically significant feature of the world", and their presentation typically exhibits what Cappelen calls "lack of a clear conclusion"; the hypothetical cases present real puzzles that subsequent philosophical analysis is supposed to solve. As such, they are often starting points of philosophical theory building.

A good example for this function is provided by a certain interpretation of the trolley case pair, as one often finds it in ethics textbooks. One case of that pair is the so-called "standard trolley problem". In the standard trolley problem, you are standing by a railroad track when you notice that an uncontrolled

trolley is rolling down the track, heading for a group of five railroad workers. You know that they will all be killed if the trolley continues on its path. The only thing you can do to prevent these five deaths is to throw a switch that will divert the trolley onto a side track, where it will kill only one person. When considering what you should do in such case, it seems that you should throw the switch and save the lives of five for the life of one. The other case of the pair, the "footbridge case", is very similar. Again, the trolley is about to kill five people. This time, however, you are not standing near the track, but on a footbridge above the track. Also, this time you cannot divert the trolley by throwing a switch. You can only stop the trolley by putting something in its way that is heavy enough. You consider throwing yourself in front of the trolley, but you know that you are not heavy enough to stop it. Standing next to you, however, is a very large stranger. The only way you can stop the trolley killing the five people is by pushing this large stranger off the footbridge, in front of the trolley. If you push the stranger off, he will be killed, but you will save the other five. Again, considering what you should do, this time it seems wrong to push the stranger off the bridge even though, as before, you would save the lives of five for the life of one.

In Thomson's (1976) presentation of the two cases, her intention was to provoke a theoretical analysis of why it is that we react differently to the two cases although they seem on a par with respect to their morally relevant aspects (it seems that in both cases you need to decide whether you should kill one in order to save the lives of five by active intervention). Puzzle cases are cases in which it is either not clear what we should say about the case, or where our pretheoretic judgement about the case seems incoherent (as in the trolley case), and they are typically intended to provoke or motivate a theoretical analysis.

Since the intention is to bring a certain type of case to the attention of philosophers, the case described must be relevant: It must be a *possible* case. In this way puzzle cases have to satisfy more constraints than mere illustrations. However, a puzzle case, just like an illustration case, does not have to elicit any clear intuitive judgement. They are puzzles – they will provoke theoretical analysis precisely because it is not clear what to say about them.

3.4 Counterexamples

Certainly, the most prominent kind of thought experiments is thought experiments that are intended to function as counterexamples. Some philosophers have even seen that as a defining feature of thought experiments. These thought experiments are intended to lead to justified belief revision in the addressee. Because of that, they need to satisfy much stricter constraints. Again, the case described must be a possible case, and it must be possible in

the right way, the description of the scenario to be evaluated should not beg the question against the theory under attack, etc. We will look at this function in detail in section 4.

3.5 Do Thought Experiments Have a Life of Their Own?

What I want to argue in this section is that even if all thought experiments that are used as counterexamples would involve intuitions, it is not clear that that function (being a counterexample) is very widespread within philosophy. As the overview that I just gave shows, not every hypothetical case that plays some role in a philosophical text is a thought experiment (some are merely analogies), and even those that count as thought experiments (according to our definition) do not all lead to the kind of worries that, for example, experimental philosophers seem to have. Thought experiments that are used as illustrations and thought experiments that play the role of puzzle cases may or may not involve intuitive judgements. When they do, the reliability of these judgements is irrelevant for the satisfaction conditions of their function.

One way to understand the argumentation in (Cappelen, 2012) is to read it as an attempt to show that most thought experiments that people have sorted in the category "counterexample" do in fact belong into the category "puzzle case". Consequently, some of the discussion of Cappelen's arguments was then over whether he looked at the real cases and whether he sorted the cases he looked at into the right categories (cf. Chalmers, 2014).

I do not find it surprising that there is controversy over how to interpret certain thought experiments and that people come to hold different views about the function of a thought experiment depending on whether they read the original text in which the thought experiment was first put forward, or a later textbook version of it. I think this is partly (but largely) to be explained by the fact that thought experiments can have lives of their own (as Ian Hacking once prominently denied, cf. Hacking, 1992). What does that mean?

A certain hypothetical case *C*, invented by philosopher *P*, might start its philosophical career as a puzzle case, either only in the mind of *P* or perhaps also in her writing. Maybe it is a puzzle case for *P*, because she realized that what she intuitively or pretheoretically wants to say about the case is different from what she thinks she *should* say on the basis of theoretical analysis, AuQ6 applying the current theory *T* to *C*. Other considerations and further reflection lead *P* eventually to a view about *C* that agrees with her pretheoretic evaluation. *T* is mistaken and has to be revised. *P* herself might now see *C* as a counterexample to *T*. *P* now presents a new theory, *T** and presents it in writing. Explaining what the difference is between *T** and *T*, she uses *C* in order to show how the content of *T** is different from the content of *T* (just as we did above with our example of the disappearing sun).

C can have these different functions for *P*; *C* can have these functions, of course, also for the readers of *P*'s writings. Perhaps they find *C* to be a challenge, perhaps they are immediately convinced that *C* refutes *T*, perhaps they already were convinced that the old theory is mistaken and see *C* as a brilliant illustration of what is wrong with it. Or they were already tacit believers of *T** and see *C* as an illustration of its merits. Accordingly, the function that *C* has in (some of) the writings of *P* and the function it has in textbooks or in the accounts of other philosophers might well differ. Even if there is a fact of the matter of how *P* intended *C* to function in a specific text[15], it might well be that readers with different backgrounds interpret the dialectical function of *C* differently.

This complicates matters for quasi-empirical studies such as Cappelen's, which want to establish that thought experiments in most/many/all cases have this or that function (of course, the same holds for those philosophers that Cappelen is attacking, who make equally general claims about the proper interpretation of thought experiments).

4 THREE MODELS FOR COUNTEREXAMPLES

OK, so the fact that we find considerations of hypothetical cases everywhere in philosophy does not by itself establish that there is a method common to much of (analytic) philosophy. Also, not every time that *we* think that a hypothetical case serves a certain dialectical function (such as that of a counterexample) does the case serve that function (for the author, or other readers). Hence initial appearances might well be quite misleading and it might be that there is not as much methodological unity to philosophy after all.

You might object that, still, you can find thought experiments used as counterexamples everywhere in philosophy in a very high concentration, such that even if you are mistaken about the proper function of some of them, still there must be enough to speak of a widespread method. What I want to argue in this section is that this impression is misleading too. The status of "counterexamples" differs between different philosophical "projects". The possible relevance that the reliability of intuitions could play for the reliability of the method of cases varies with that status. On several such projects, the reliability of intuitions plays no role, even if it turned out that intuitive judgements are involved in the method of cases.

15 I do not think that there always has to be a fact of the matter what communicative intentions are behind a piece of text. Especially with written text, these intentions can change in the process of writing. Just consider those cases in which you had to rationally reconstruct your own arguments.

Philosophy allows for the coexistence and the overlap of these projects in all of its subdisciplines. Therefore, any results about the reliability of intuitions are not relevant for philosophy in general, not even for specific debates within philosophy, but at best for certain philosophers in those debates, namely, those philosophers who assign a certain foundational role to intuitions. As I will argue in section 5, it seems to me that only very few philosophers actually do that. But let us first look at my argument for the project-relativity of the status of counterexamples.

4.1 Evidence

The first project is what most people writing about the method of cases and criticizing the reliability of intuitions seem to have in mind, the *Evidence Model*. Here, the aim of philosophical research is typically considered to be an inquiry into certain matters of fact, which are themselves "intuition-independent". With the latter I mean that the facts in question are neither constituted by intuitions, nor by the psychological basis for these intuitions. For example, on a current understanding of metaphysics, metaphysicians try to find out what "grounds" what. They assume that there is a matter of fact of what grounds what, and that fact is neither constituted by our intuitions, nor by whatever psychological mechanism gives rise to our having these intuitions. If, on such a conception of the subject matter of philosophical inquiry, you take your intuitive evaluation of a hypothetical case as evidence about that subject matter, then you follow a certain type of philosophical project, which I call the "Evidence Model". You might or might not have a story about why your intuitive evaluation of such a hypothetical case is evidence about those facts that you intend to investigate. In either case, learning that the mere intuitive evaluation of hypothetical cases varies much between persons, is *prima facie* bad news for you. However, even on this model, there is some room for manoeuvre (and we will discuss this in more detail in section 5).

4.2 Reflective Equilibrium

However, not everyone conceives of the subject matter of philosophy in that way. Resnik (1999), for example, holds that there is no fact of the matter of whether something is a logical truth, others are non-factualists about moral truth, or ontology, etc. Still, these philosophers might nevertheless follow a methodology which involves the intuitive consideration of hypothetical cases and the use of such cases – in some sense – as counterexamples to proposed theories. Nelson Goodman (1955) described this methodology as that of finding a *reflective equilibrium* between our intuitive evaluation of cases and our endorsed general principles. In that process, sometimes general principles are

given up, because we decide to consider an intuitive evaluation of a hypothetical case to be worth preserving, sometimes the general principle trumps our intuitive evaluation, perhaps because it is better entrenched. In the former case, the hypothetical case works as a counterexample against the proposed theory.

However, on this project our intuitive evaluations are not evidence about a further subject matter that this methodology seeks to investigate. There are no facts of the matter about these things. Learning that intuitions are prone to influences that we do not have under control or that they vary much between persons is interesting but neither very surprising nor very worrying. We thought we need to regiment those intuitive evaluations anyway.

4.3 Constitution

Still, your account of what you are up to in philosophy might differ from both of these. Perhaps you do believe that there are facts of the matter of what philosophers are interested in, but these facts are not intuition-independent. In his (2007) Alvin Goldman argues that philosophers are interested in investigating concepts in a personal psychological sense, and that these are, at the same time, the psychological basis for our dispositions to evaluate hypothetical cases intuitively in certain ways (Goldman, 2007, p. 15):

> It's part of the nature of concepts (in the personal psychological sense) that possessing a concept tends to give rise to beliefs and intuitions that accord with the contents of the concept. If the content of someone's concept *F* implies that *F* does (doesn't) apply to example *x*, then that person is disposed to intuit that *F* applies (doesn't apply) to *x* when the issue is raised in his mind.

Still, on this model, your intuitive evaluation of a hypothetical case can serve as a counterexample to a proposed theory about what the psychological concept in question is. But, again, the relevance of, for example, interpersonal variation in application intuitions is limited. If you are first of all interested in personal psychological concepts, then inter-personal differences in such concepts might be interesting, but not a threat of any kind to your methodology. If you connect further aims with your study of personal psychological concepts (for example, if your ultimate aim is to study the content of concepts in a certain population), then too much interpersonal variation might either lead you to distinguish between competence and performance (if there is an available causal explanation for the interpersonal differences in performance) or it might convince you that your project is pointless. But then this is because your assumptions about the subject matter were mistaken (the subject matter is too messy to allow systematic theorizing), but not because intuitions turned out to be too unreliable evidence for that subject matter.

These are just three stereotypical ways in which philosophers might conceive of the project that they are engaged in. Possibly there are more, and certainly, there are mixed versions of these projects. The point is that philosophers from different projects might still all participate in one and the same philosophical debate, arguing about the same theories and the same proposed counterexamples, albeit assigning very different roles to those intuitive evaluations of hypothetical cases; they might play the role of evidence about that subject matter, they might be part of the construction of the subject matter, or of the constitution of it. The reliability of intuitions, of the kind that experimental philosophers attack and that their foes defend, is only of relevance to some of the participants in these debate and their projects.

5 INTUITIONS VERSUS COMMON GROUND

Still you might think that philosophers who follow the project characterized as the Evidence Model are clearly the majority. Few philosophers (one often reads) would nowadays declare that they are in the business of doing conceptual analysis of any kind, let alone of personal psychological concepts. Anti-realists about philosophy's subject matter are likewise the minority in many debates. So, still, is that not sufficient reason for paying close attention to the reliability of intuitions in our judgements about hypothetical cases?

The three models that I presented in the last section all had two things in common, they all used hypothetical cases as counterexamples *and* they gave intuitions a central role in the evaluation of these hypothetical cases. With the exception of the third model, in which intuitions are partly constituting the subject matter, I could have described these models without bringing in intuitions at all. We could describe the second model as one that departs from pretheoretic beliefs, and the first as one that takes judgements about hypothetical cases that are in the common ground between the parties in the philosophical debate as evidence for and against theories.

Why should philosophers do the latter? Because that way they have a better chance to convince their readers of their arguments. Why should other philosophers accept such evidence? Because they also think that the evaluation of these cases is correct. As Cappelen (2012) has shown, this description of the method of cases accounts much better for the way in which the evaluation of hypothetical cases is usually embedded in argumentations. Philosophers give *reasons* for evaluating a hypothetical case in a certain way. Reasons for the correctness of the evaluation, and typically not for the intensity or quality of any intuition. You might or might not have arrived at your evaluation of the hypothetical case purely intuitively. But it does not seem to matter for the significance of your evaluation when you did.

True, there are exceptions. Some philosophers indeed merely refer to their intuition and do not think that it is required or even possible to provide further justification of their judgement. It is a long shot to argue that philosophers who follow that program are the actual majority. If they are not then it might seem as if the Evidence Model is widespread in philosophy, while in fact intuitions play no terribly significant evidential role.

6 CONCLUSION

I sketched different philosophical projects and argued that the question of whether intuitions are reliable is not of relevance for the majority of these projects. I also argued that the evidence (for example, the evidence collected in Cappelen (2012)) suggests that this is not only true for the majority of logically possible philosophical projects, but true for the majority of the actually implemented ones.

So, the methods which can – at an abstract level – all be made to fit under the description "using hypothetical cases as counterexamples" are in fact too diverse to allow for an interesting, general metaphilosophical analysis. This is a descriptive metaphilosophical result. The result is based on a rather liberal attitude towards those different projects. Perhaps many of the projects that I identified are ill-conceived. Even though I found philosophers who seem to endorse them, perhaps they do not in fact follow them (on rational reconstruction of their work), but just fell prey to a metaphilosophical confusion. Once the ill-conceived projects are eliminated, could it not be the case that interesting metaphilosophical claims can be made about the remaining scene?

That is probably so. I even believe that it is definitely so, since I believe that I have a somewhat clear idea of what I do when I do philosophy, and also believe that this is a better way of doing it then all the alternatives. But the reasons for following one project rather than another are typically not just matters of internal consistency or feasibility in principle. They are also matters of personal interest and subjective pragmatic judgement. Perhaps the role of metaphilosophy could be to provoke reflection on the relevant issues in order to increase the number of philosophers who work within a well-conceived program. But that will still probably not result in a convergence of these programs. It is just too obvious that the standards for what counts as "philosophically interesting" are very different even between otherwise very like-minded philosophers.

So philosophy is just stuck with methodological pluralism. But, as I said before, this is to be expected and it is not actually detrimental for philosophy. It is to be expected, because in philosophy everything is always up for grabs. The idea that philosophy as a whole should follow something like a

theoretically defined paradigm seems to go against the very nature of the subject. That is why we should not expect convergence of philosophical projects. It does not seem to be a problem either, since, as I also indicated before, even if philosophers engaged in different projects might interpret the role and value of judgements about hypothetical cases differently, on none of the accounts that we looked at do hypothetical cases come out being irrelevant. Philosophers of all projects sketched can and do participate in the same debates about the same theories and the same proposed counterexamples. Moreover, they even manage to sometimes reach an agreement about when a theory should be abandoned in light of such counterexample.

REFERENCES

Cappelen, H. (2012). *Philosophy without intuitions*. Oxford: Oxford University Press.

Carnap, R. (1967). *Logical foundations of probability*. Chicago, IL: University of Chicago Press.

Chalmers, D. J. (2014). Intuitions in philosophy: A minimal defense. *Philosophical Studies*, *171*, 535–44.

Goldman, A. (2007). Philosophical intuitions: Their target, their source, and their epistemic status. *Grazer Philosophische Studien*, *74*(1), 1–26. https://doi.org/10.1163/9789401204651_002

Goodman, N. (1955). *Fact, fiction, & forecast*. Cambridge, MA: Harvard University Press.

Hacking, I. (1992). Do thought experiments have a life of their own? Comments on James Brown, Nancy Nersessian and David Gooding. *Philosophy of Science*, *2*, 302–8.

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford: Oxford University Press.

AuQ7 Pust, J. (2016). Intuition. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

Resnik, M. D. (1999). Against logical realism. *History and philosophy of logic*, *20*(3–4), 181–94. https://doi.org/10.1080/01445349950044134

Stöltzner, M. (2004). The dynamics of thought experiments a comment on David Atkinson. In M. C. Galavotti (Ed.), *Observation and experiment in the natural and social sciences* (Vol. 237, pp. 243–58). Dordrecht: Springer.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204–17. https://doi.org/10.5840/monist197659224